RooUnfold SVD vs TUnfold

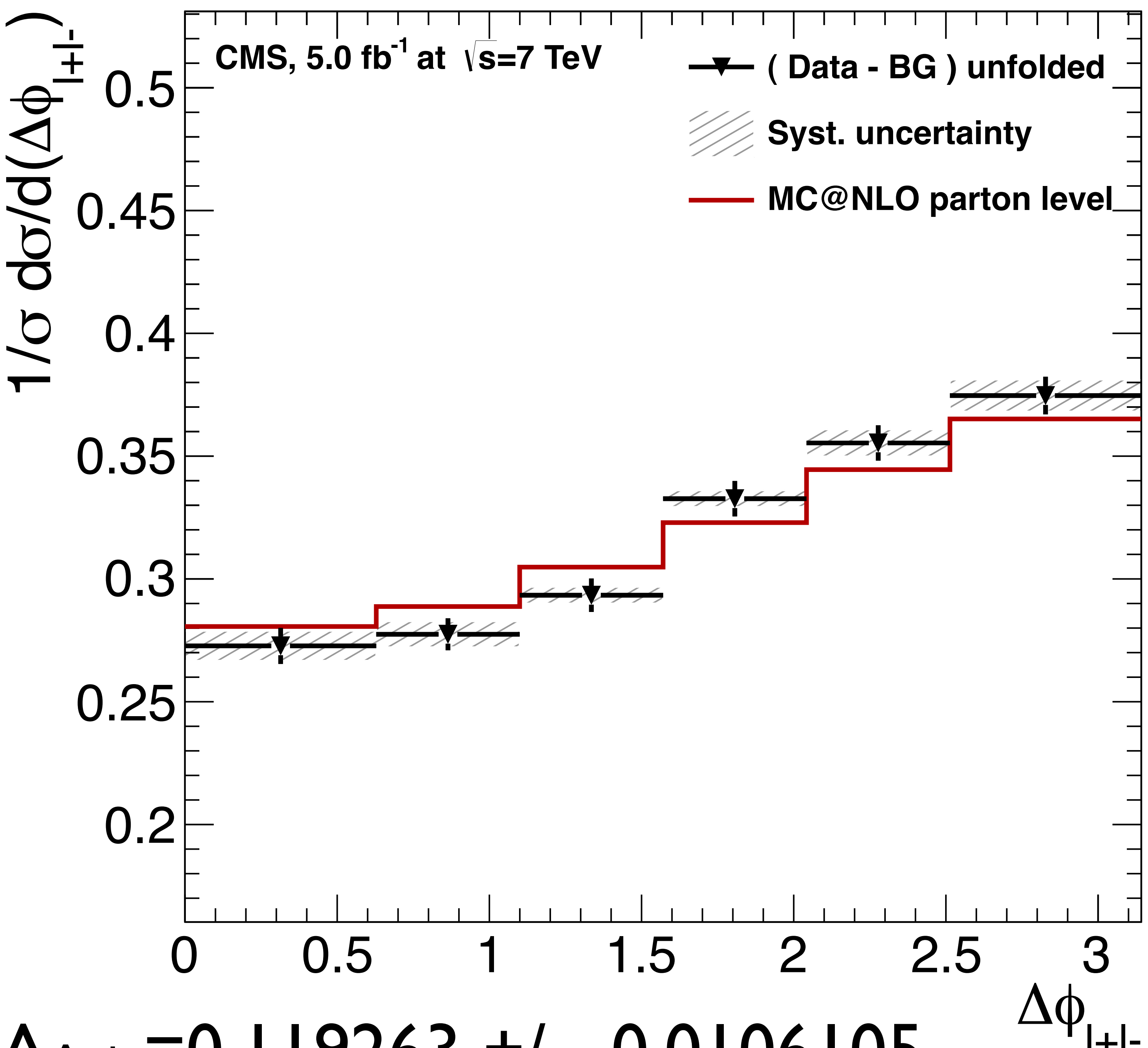
Introduction

- Tried to reproduce SVD results (as used for the papers) using TUnfold (with kRegModeCurvature)
  - used the same 6 gen bins, and 12 reco bins (each gen bin split in two)
  - ensured the bias distribution was being used correctly by setting biasScale = N_{data}/N_{MC} ~ 1 (previously it was defaulting to 0)
- The plots on the following slides show the SVD results with regularisation parameter k varied between 1 (maximal regularisation, measured = truth) and 6 (minimal regularisation), alongside the TUnfold results with tau set to approximately match the SVD results
  - we used k=3 for the paper
- Note I turned off jet smearing to make these plots, so only the purely leptonic variables exactly match the paper results (for k=3)

$A_{\Delta phi}$ results

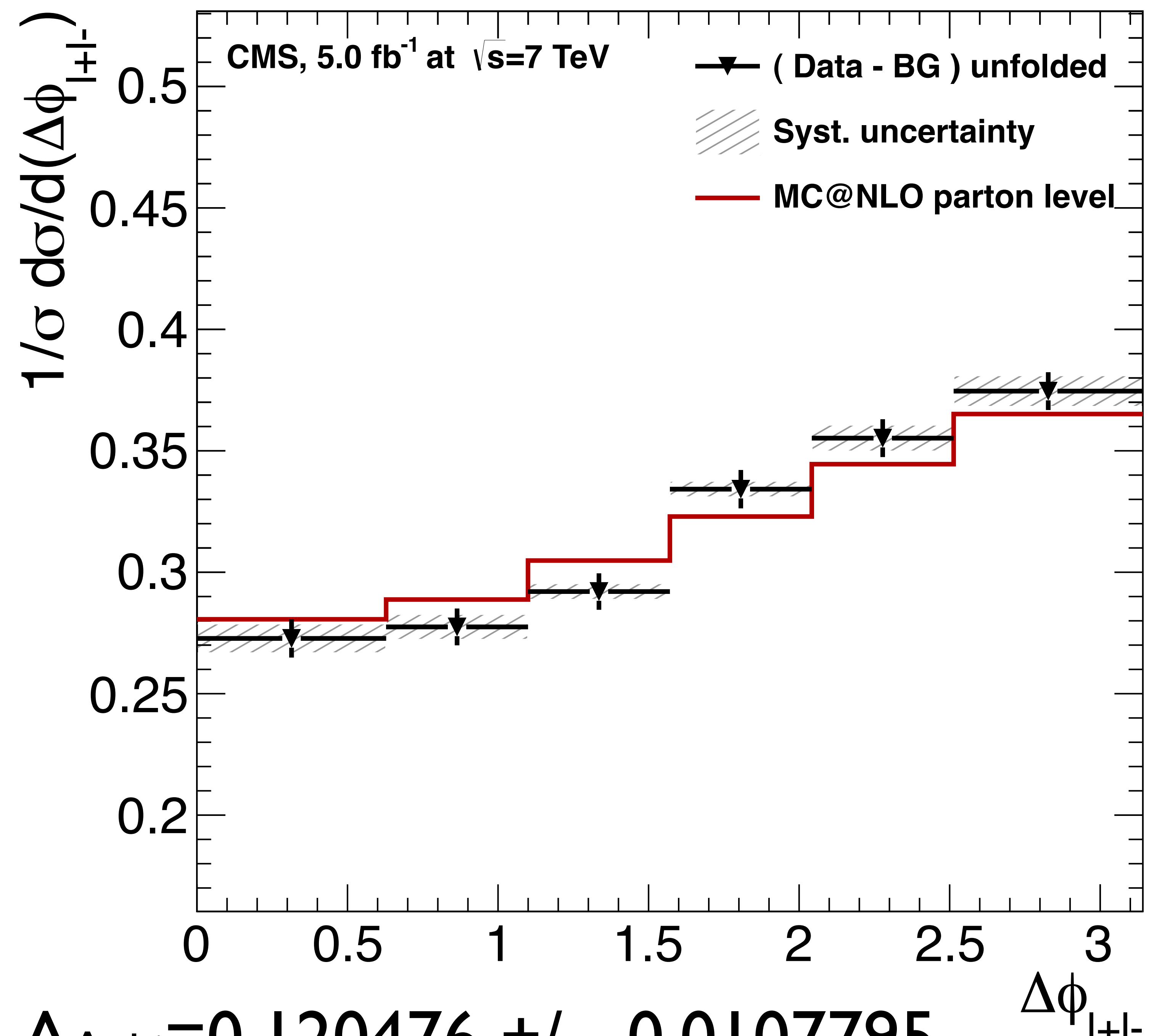
note, scanLcurve prefers tau < 0.001

SVD, k=6



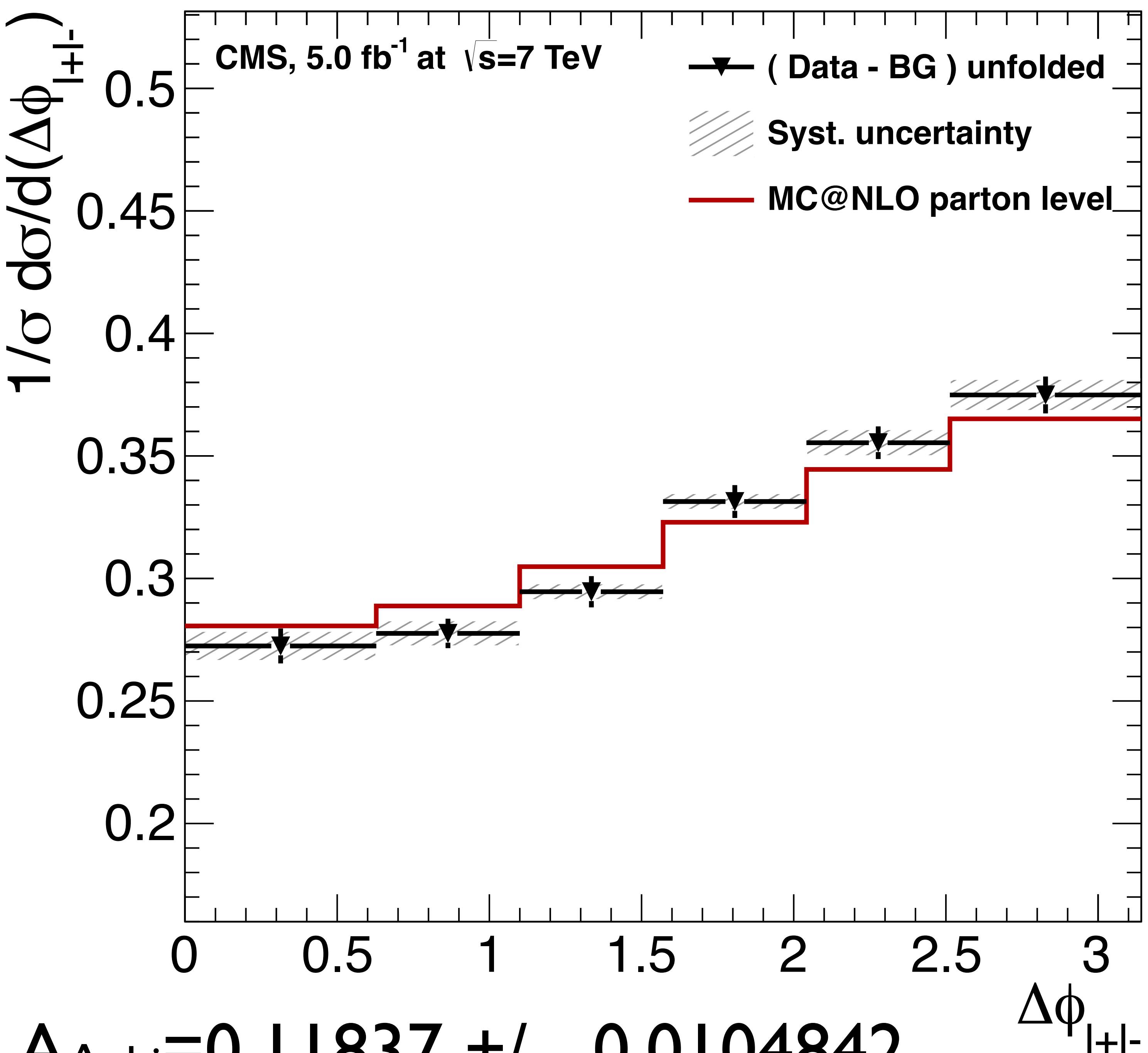
$A_{\Delta phi}$=0.119263 +/- 0.0106105

TUnfold, tau = 0.005



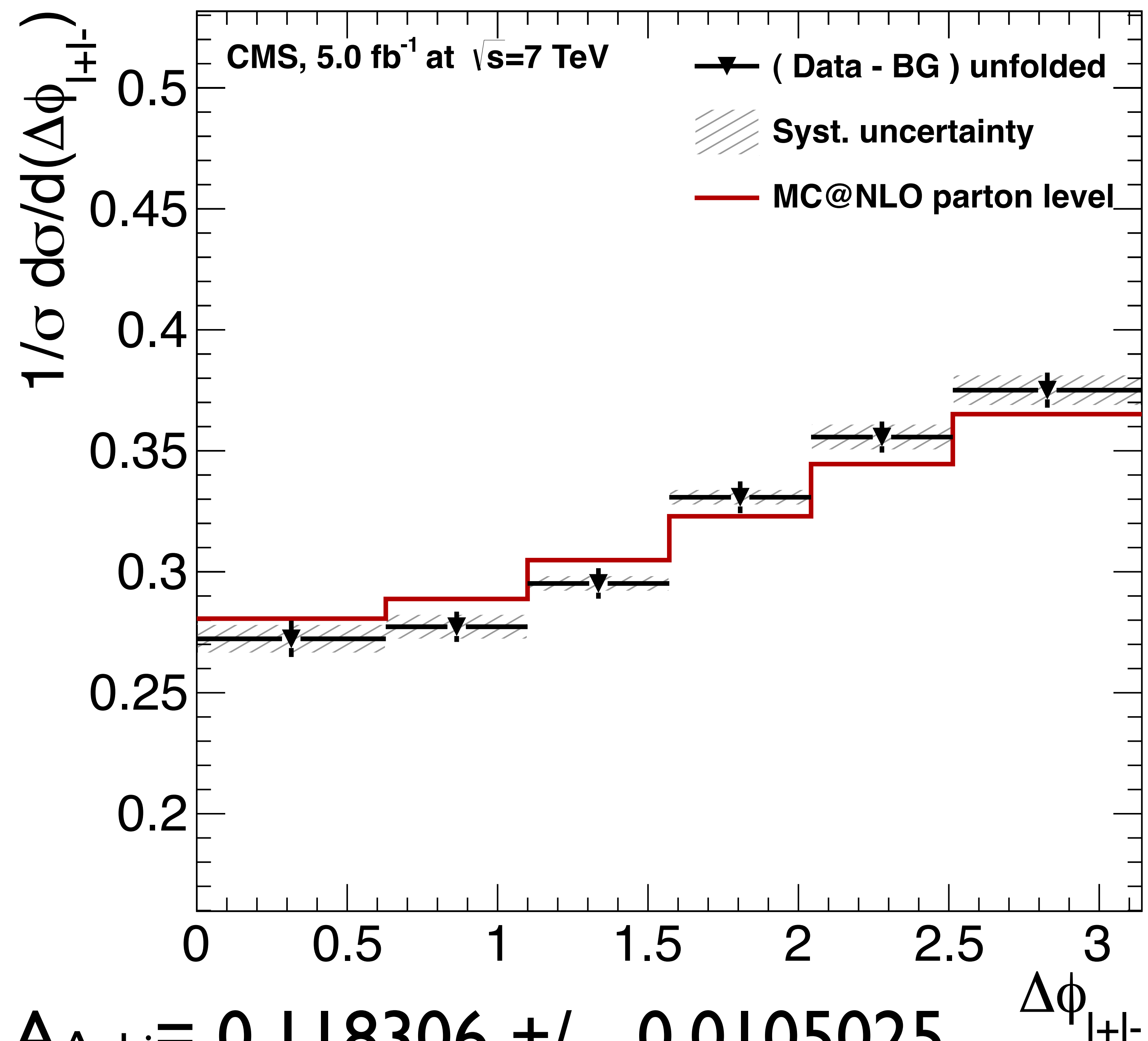
$A_{\Delta phi}$=0.120476 +/- 0.0107795

SVD, k=5



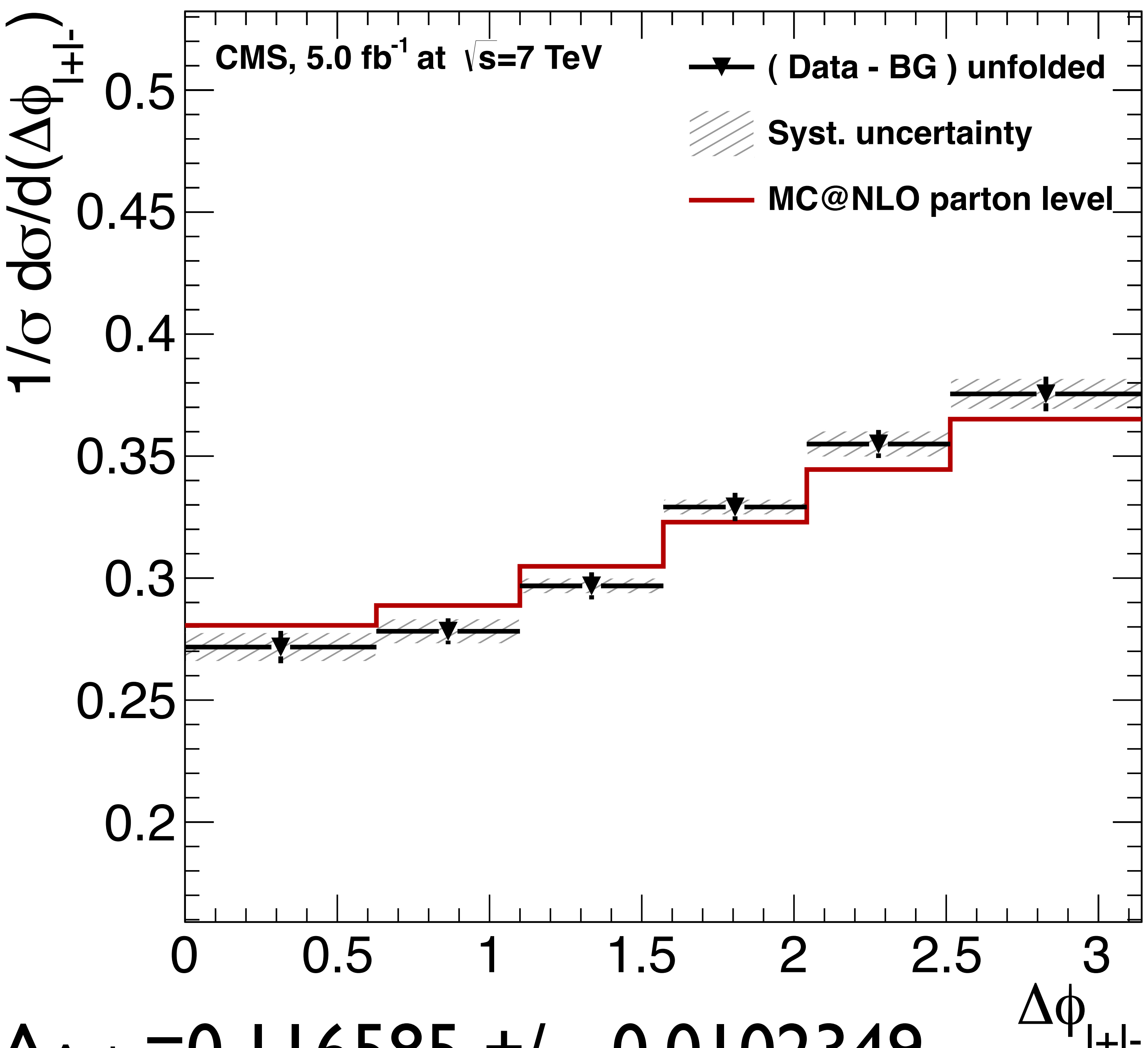
$A_{\Delta phi}$=0.11837 +/- 0.0104842

TUnfold, tau = 0.01



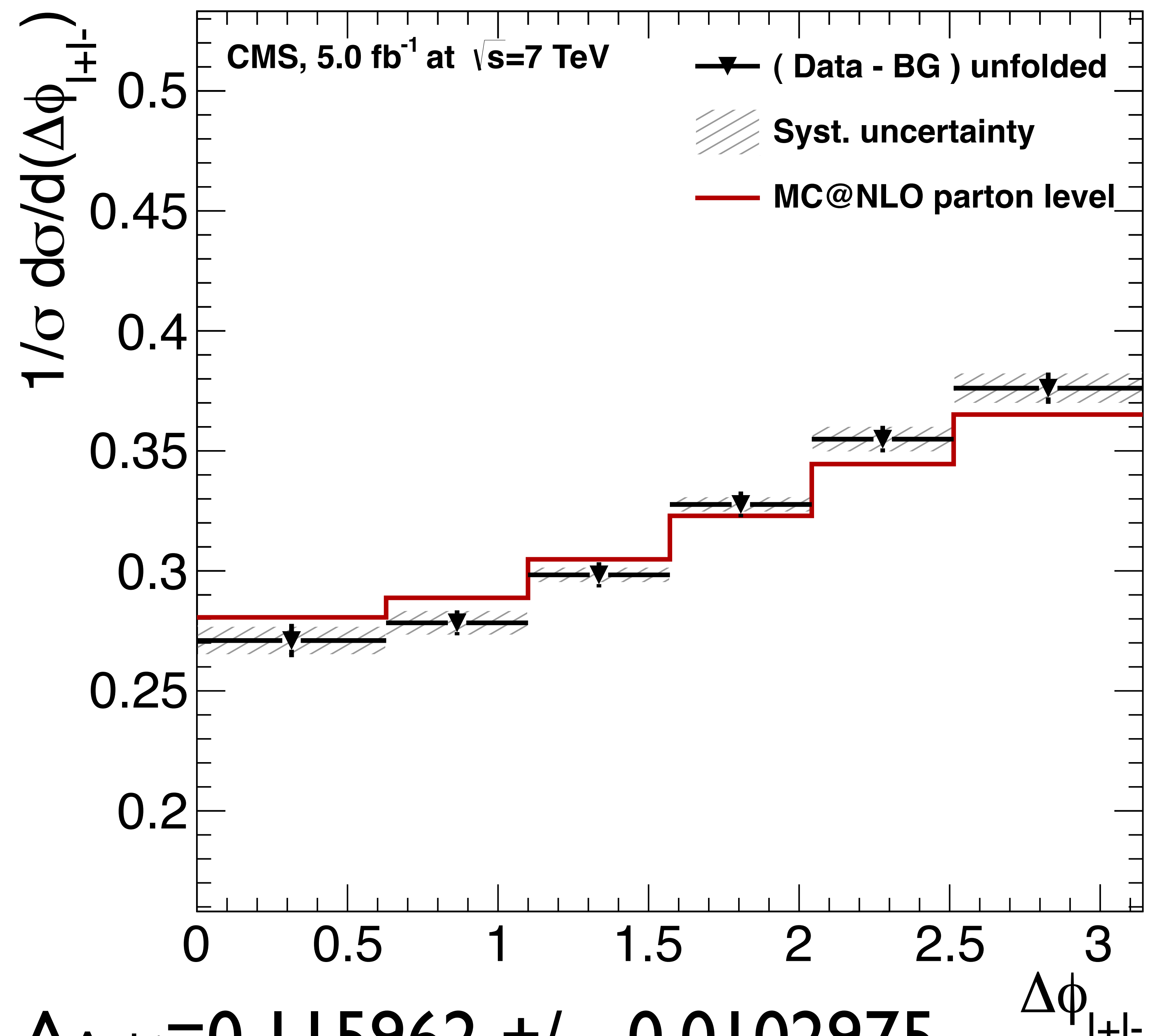
$A_{\Delta phi}$= 0.118306 +/- 0.0105025

SVD, k=4



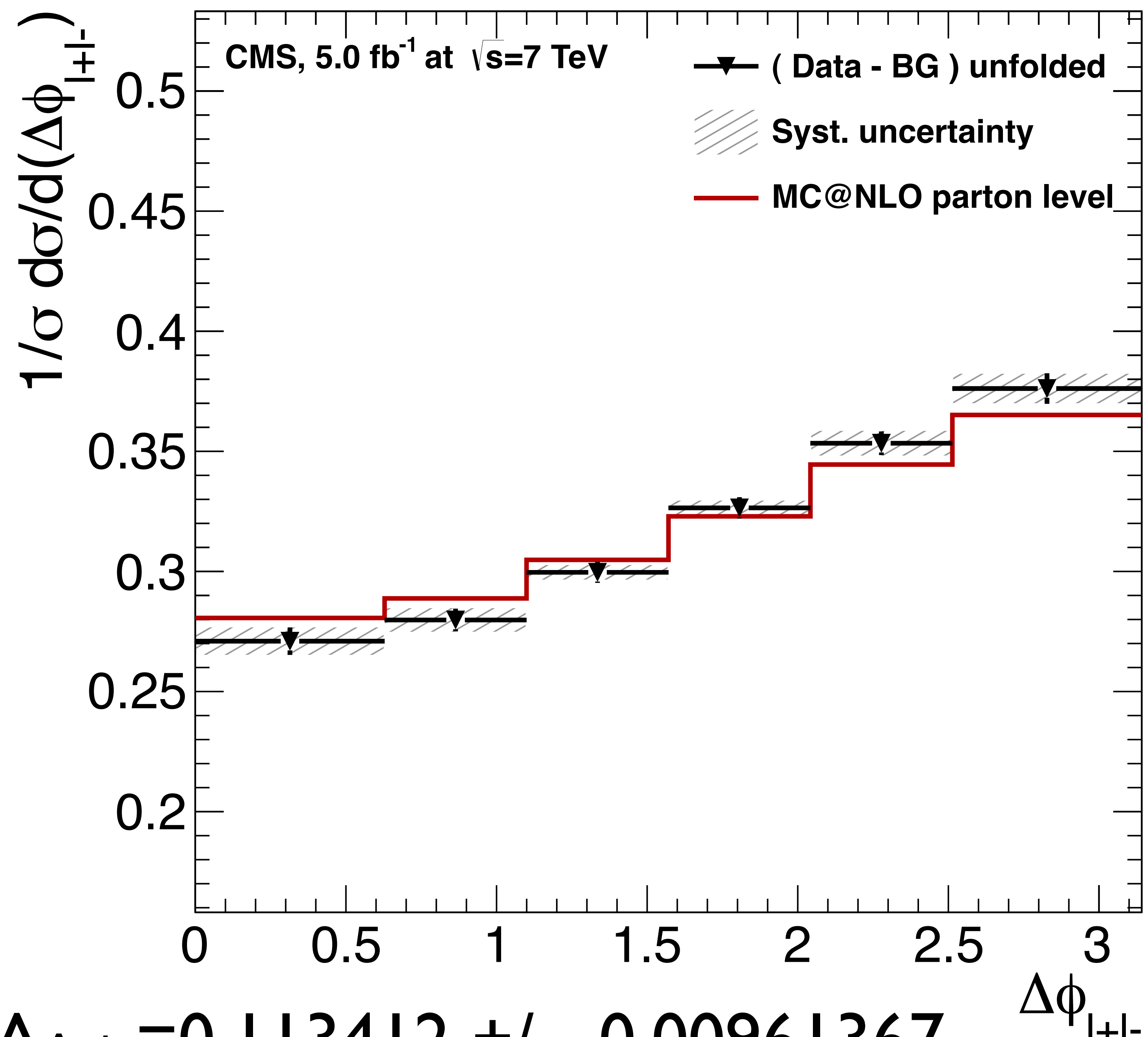
$A_{\Delta phi}$=0.116585 +/- 0.0102349

TUnfold, tau = 0.02



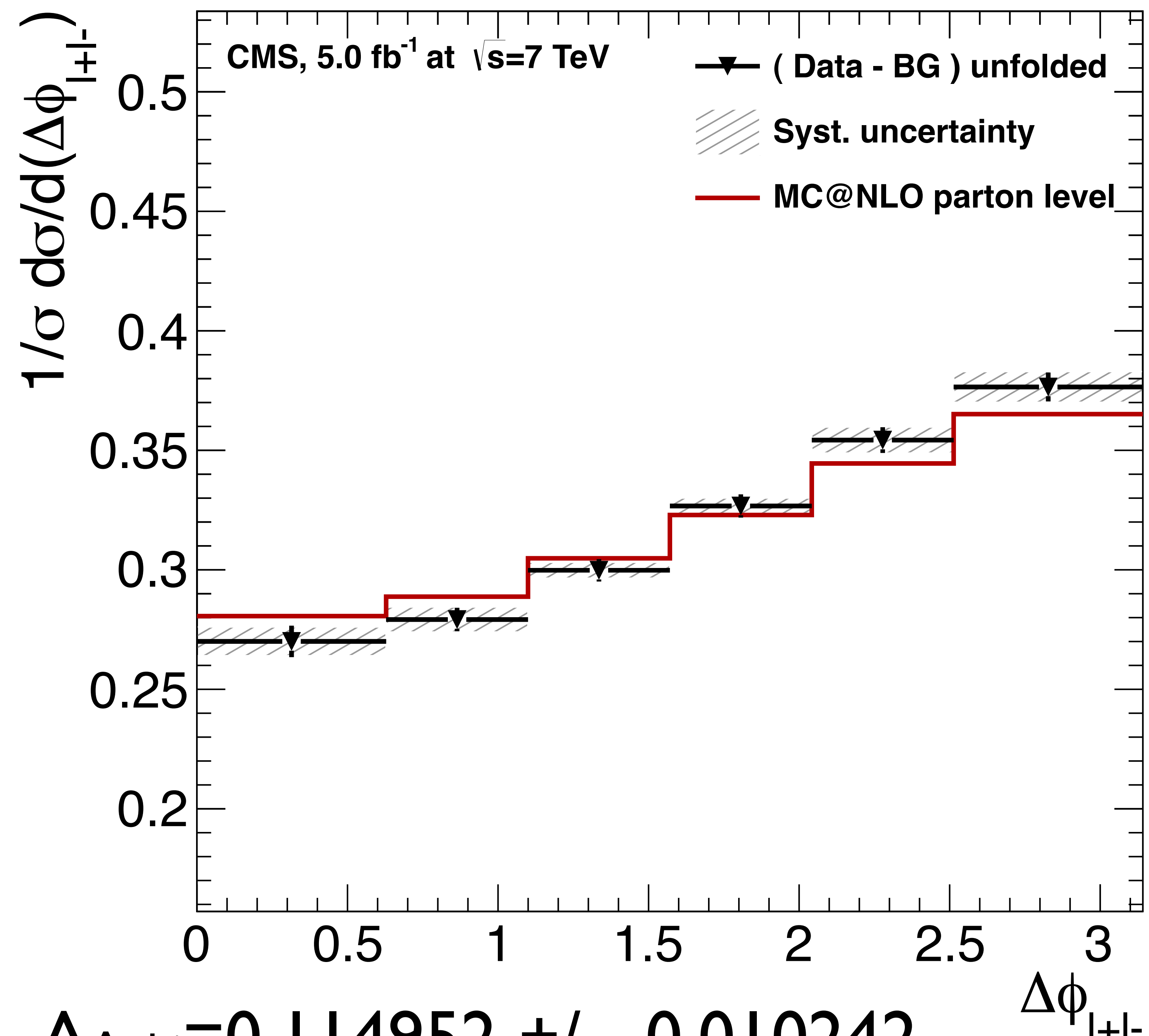
$A_{\Delta phi}$=0.115962 +/- 0.0102975

SVD, k=3



$A_{\Delta phi}$=0.113412 +/- 0.00961367

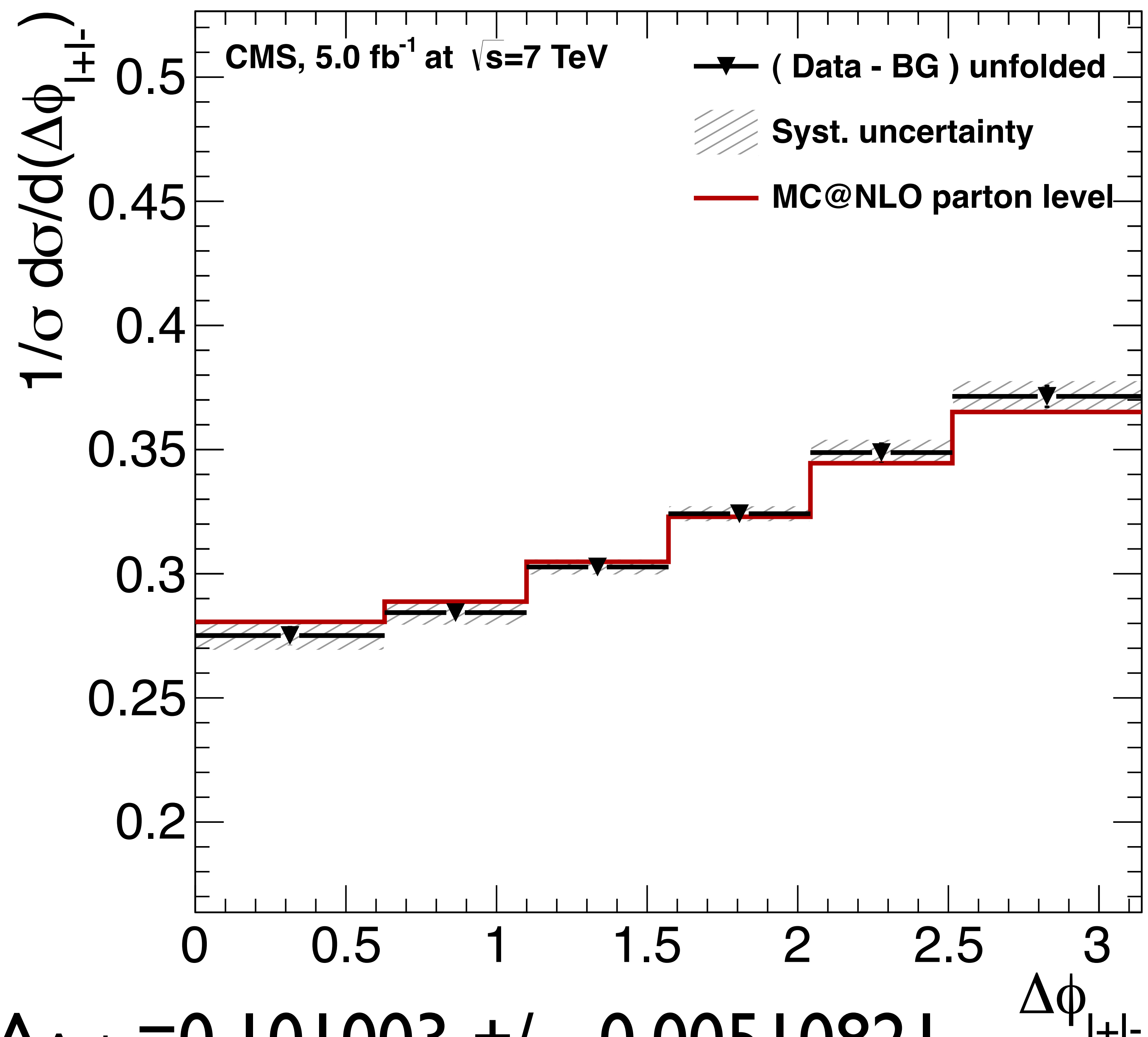
TUnfold, tau = 0.03



$A_{\Delta phi}$=0.114952 +/- 0.010242

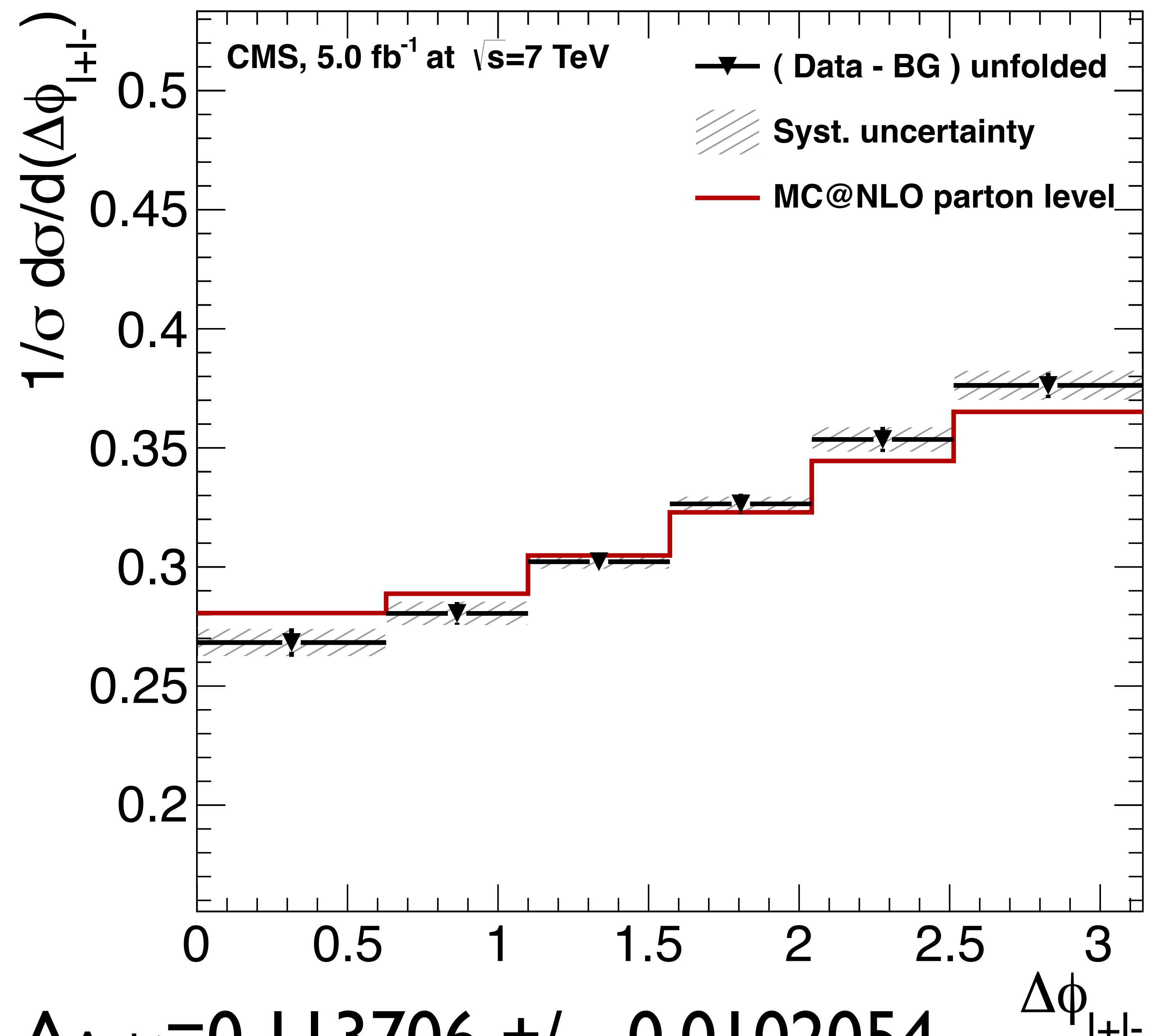
SVD clearly over-regularised here. Can't match this regularisation strength with TUnfold, even with very high tau (see next slide).

SVD, k=2



$A_{\Delta phi}$=0.101003 +/- 0.00510821

TUnfold, tau = 0.1

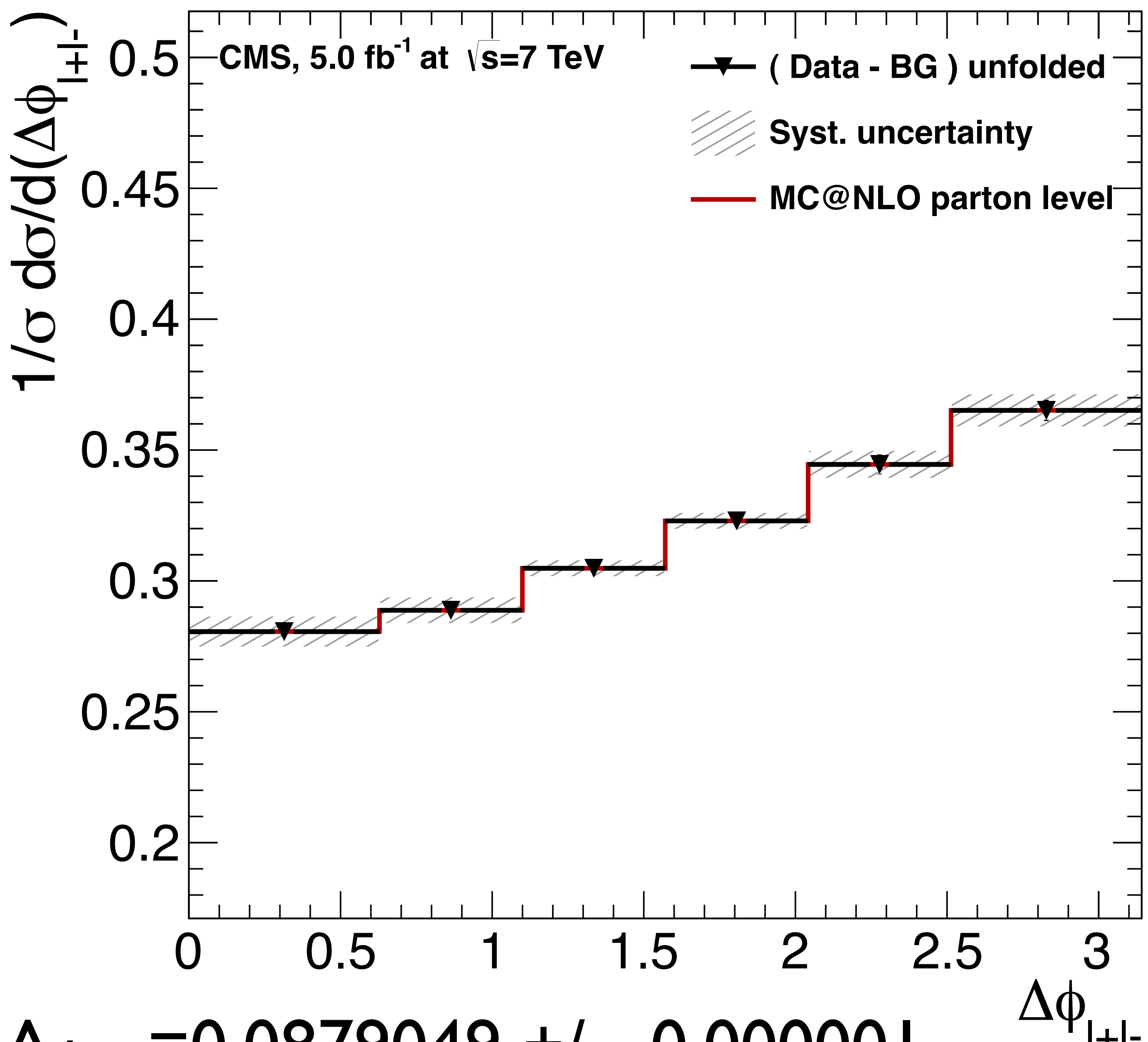


$A_{\Delta phi}$=0.113706 +/- 0.0102054

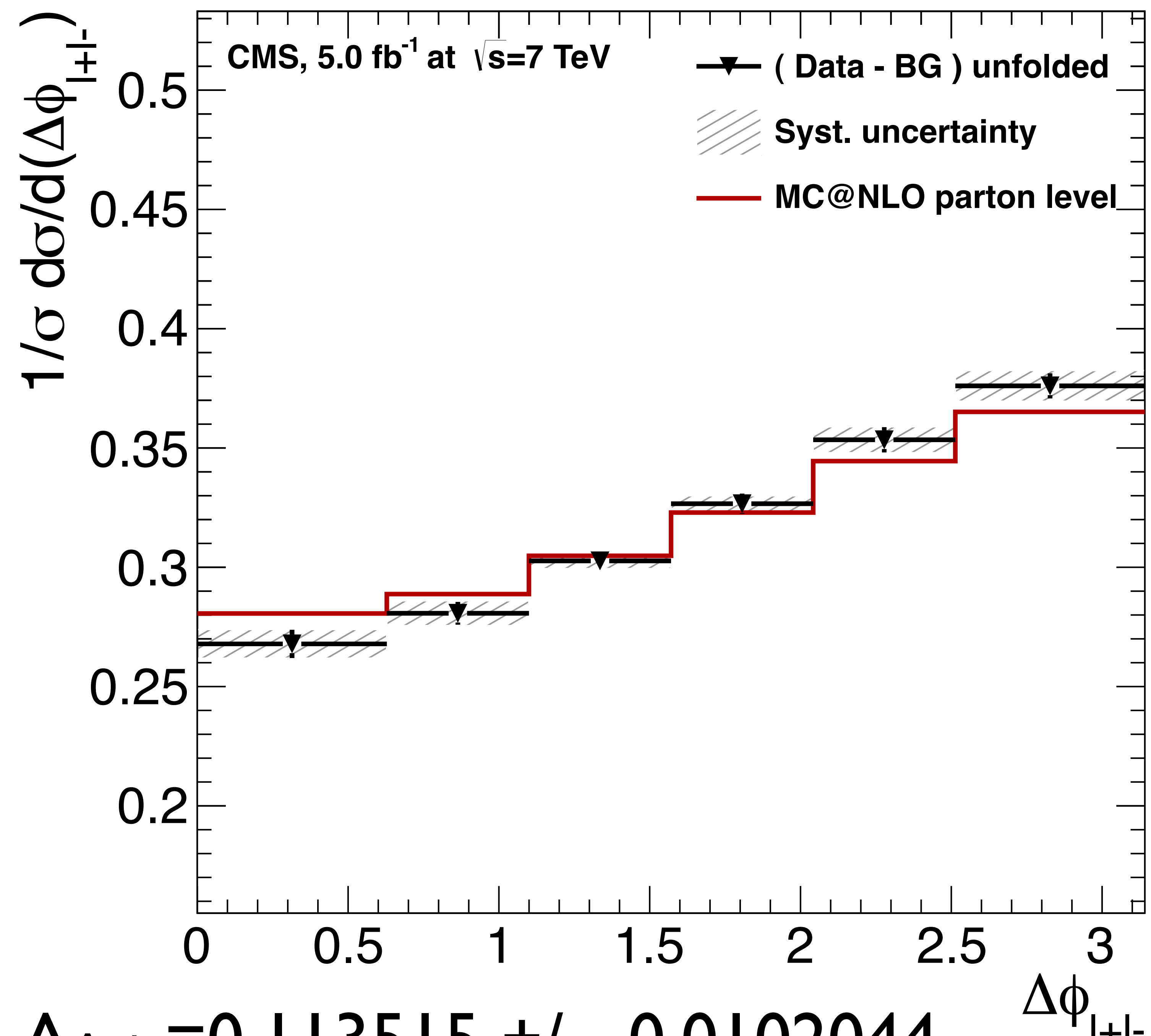
SVD, k=1

TUnfold, tau = 10

can’t match this regularisation strength, even with higher tau



$A_{\Delta phi}$=0.0879049 +/- 0.000001

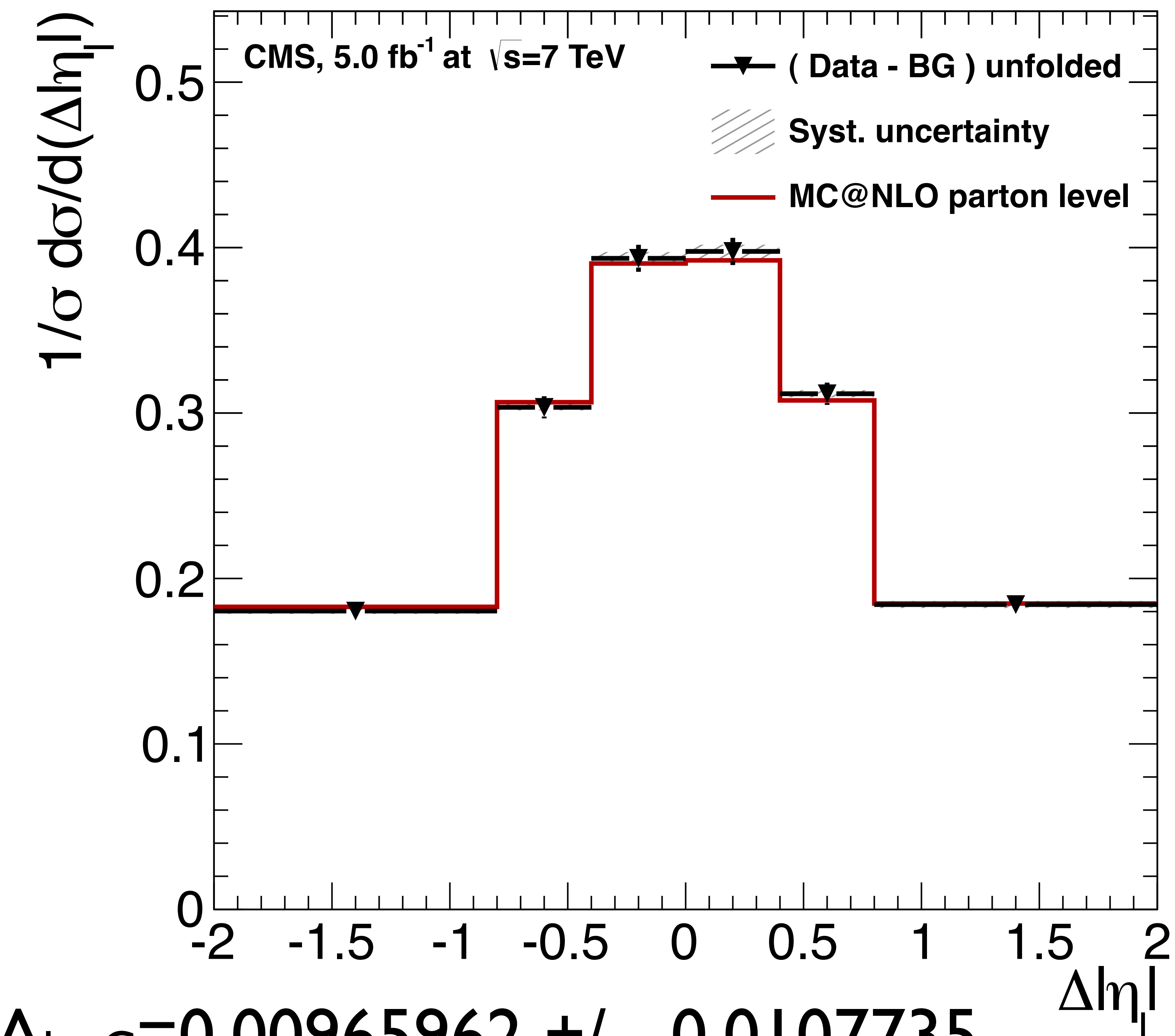


$A_{\Delta phi}$=0.113515 +/- 0.0102044

A_{lepC} results

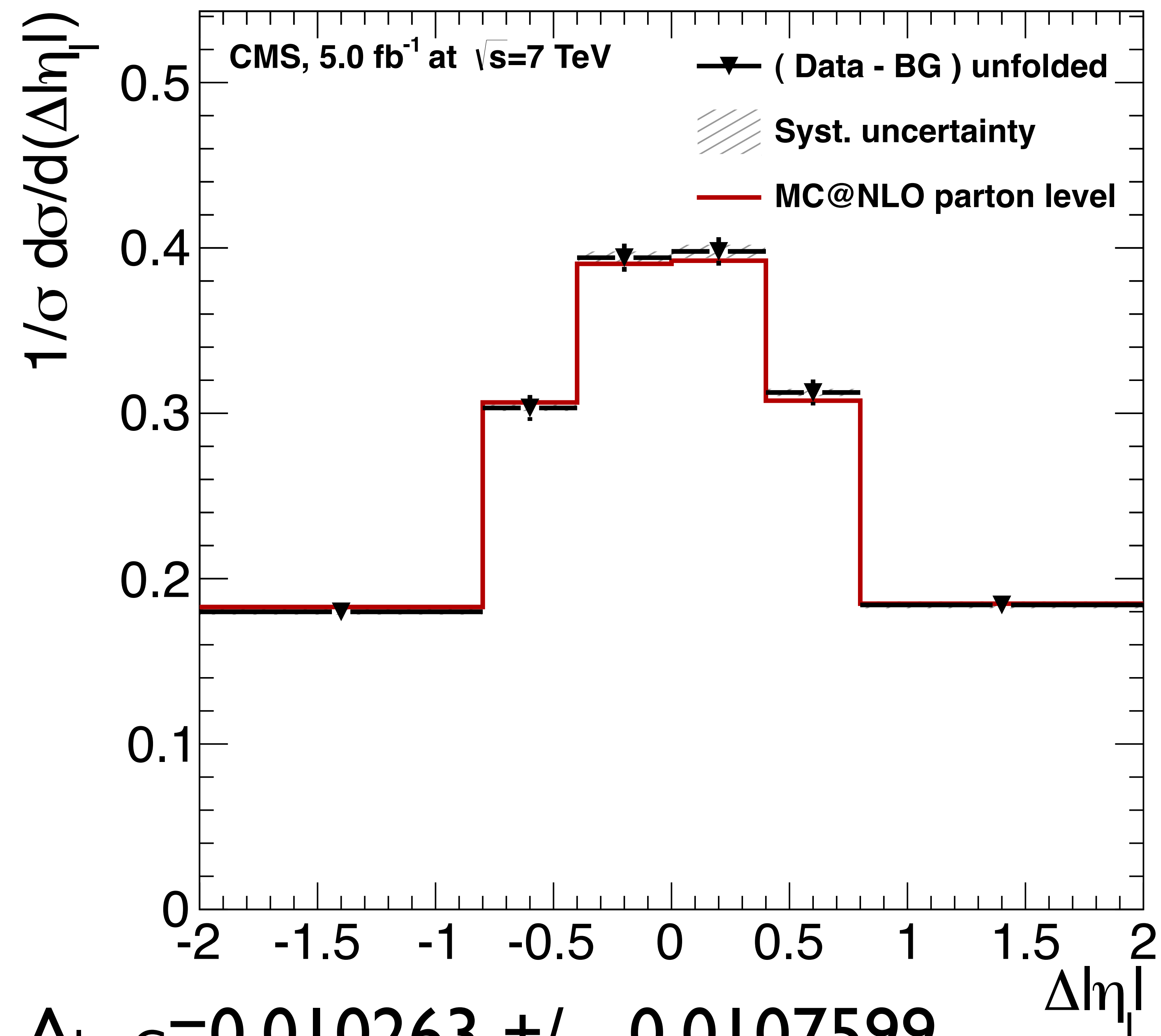
note, scanLcurve prefers tau < 0.001

SVD, k=6



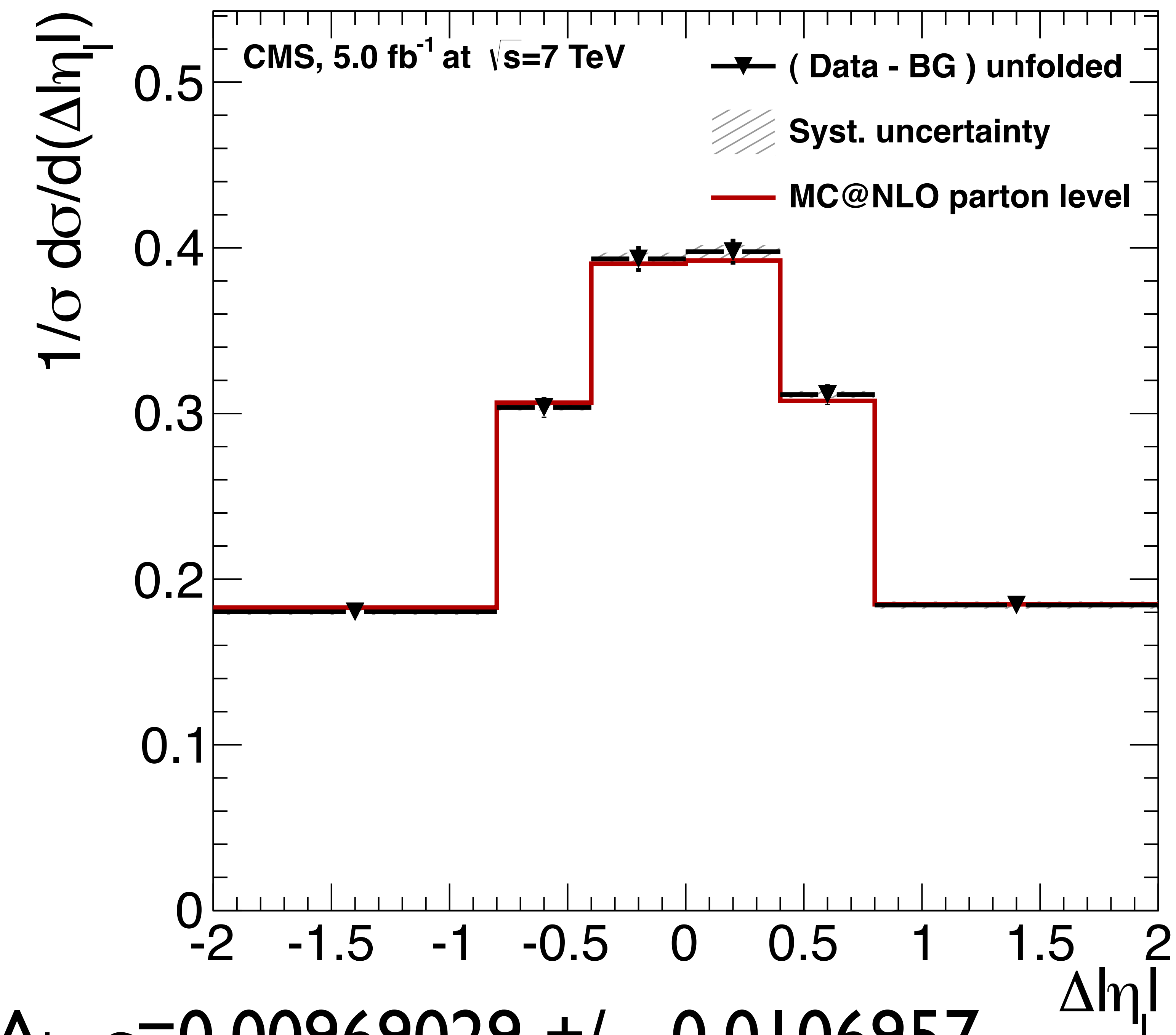
A_{lepC}=0.00965962 +/- 0.0107735

TUnfold, tau = 0.005



A_{lepC}=0.010263 +/- 0.0107599

SVD, k=5



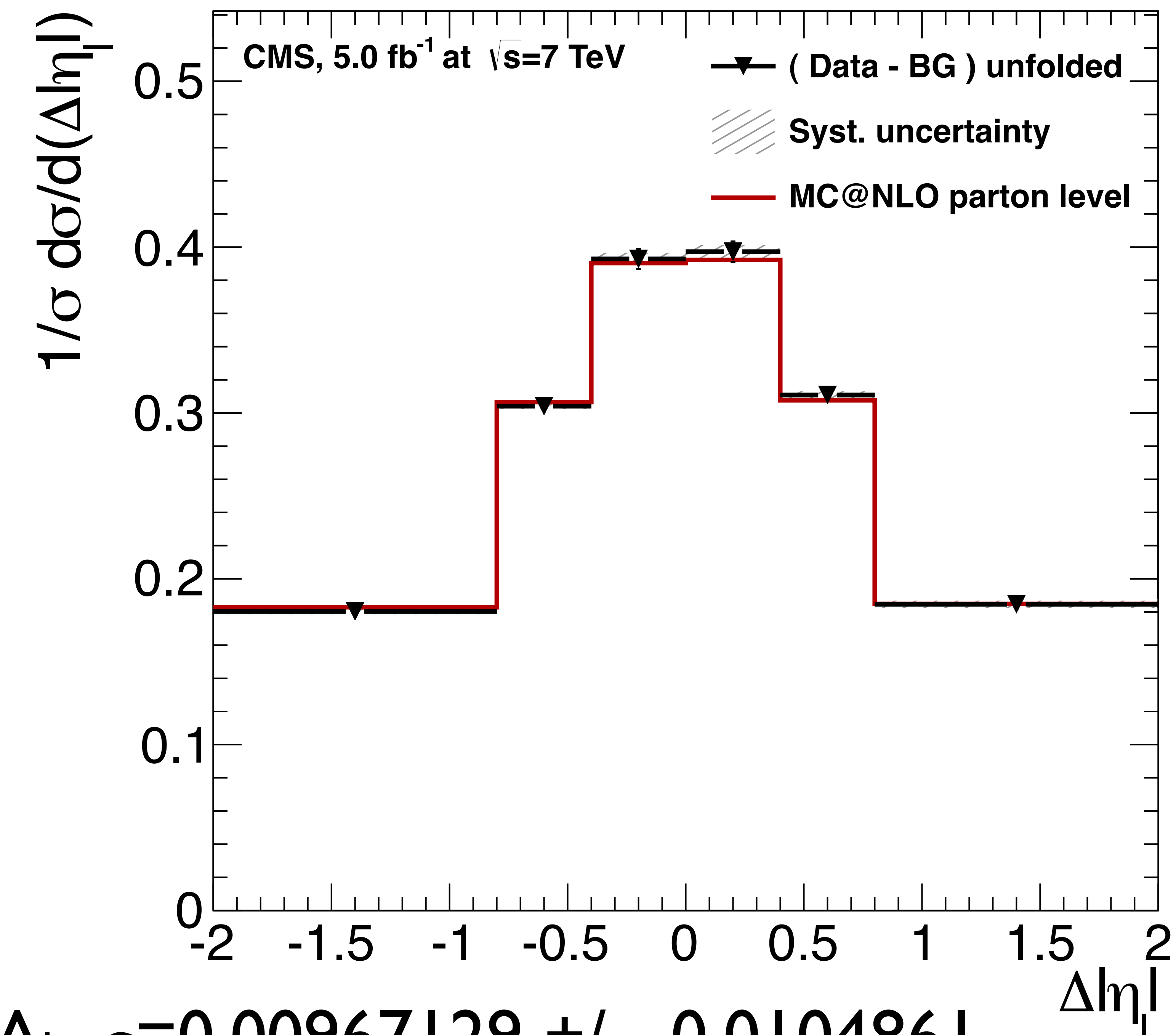
A_{lepC}=0.00969029 +/- 0.0106957

TUnfold, tau ~ 0.007

[missing plot]

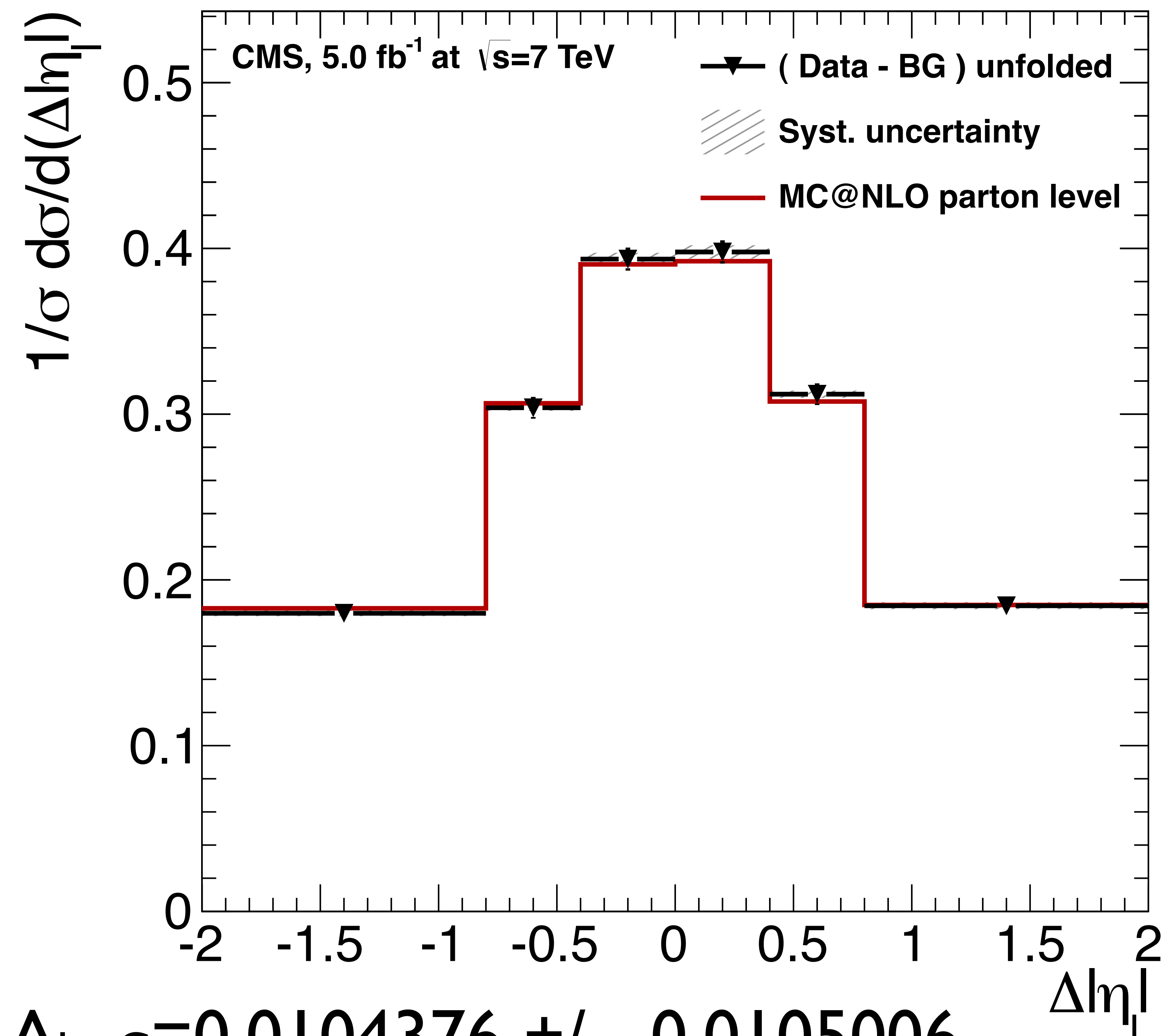
A_{lepC}=

SVD, k=4



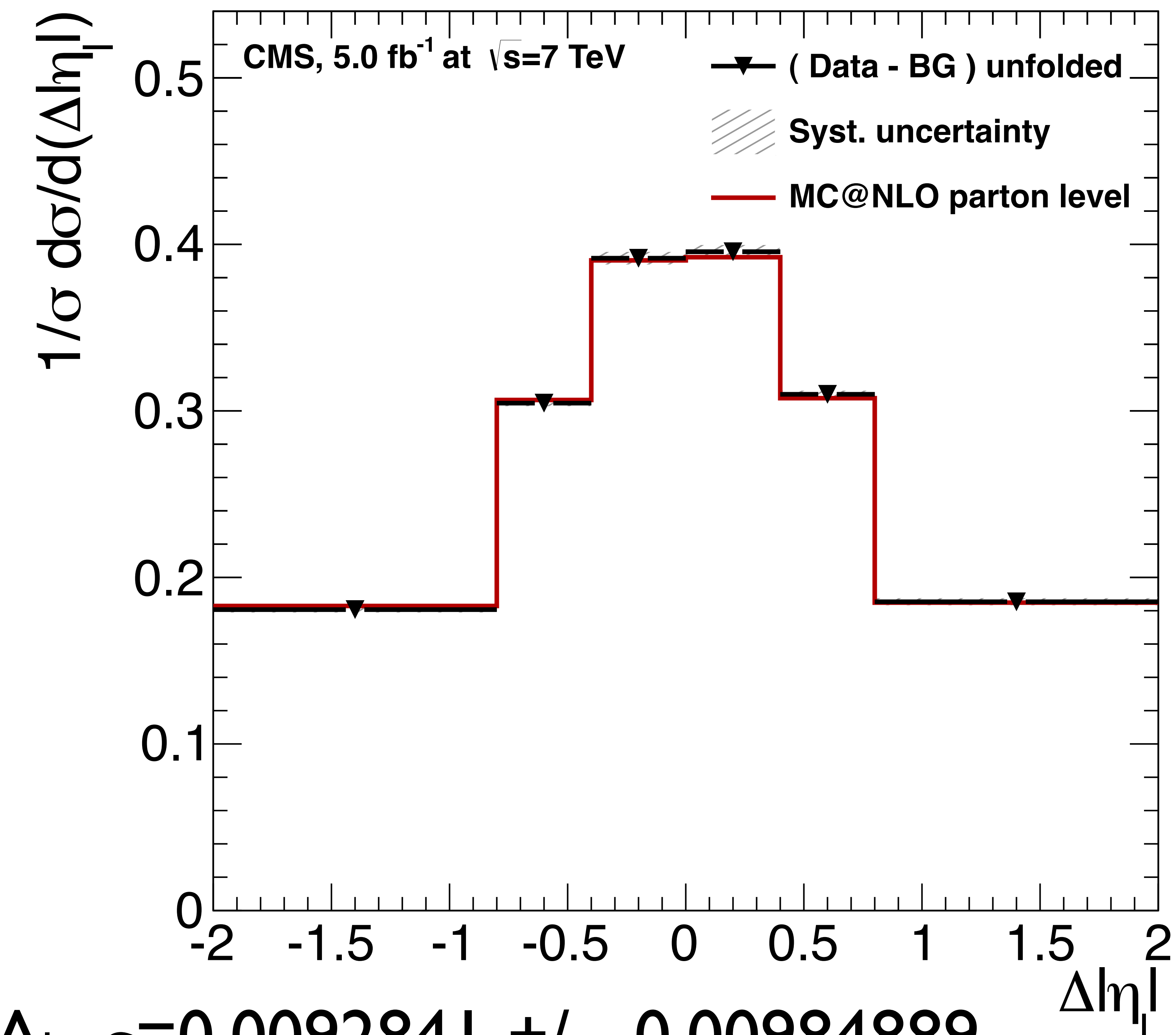
A_{lepC}=0.00967129 +/- 0.0104861

TUnfold, tau = 0.01



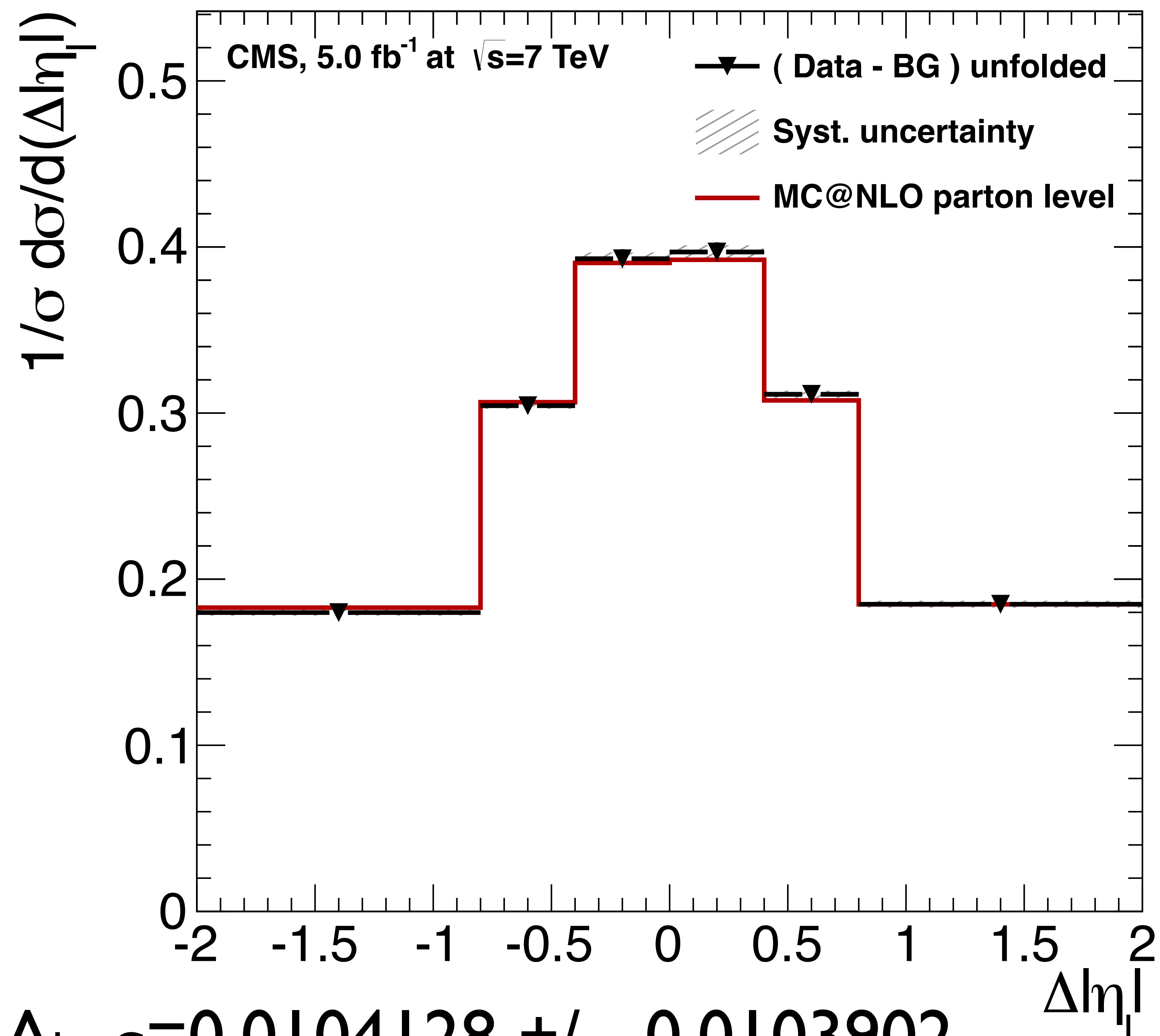
A_{lepC}=0.0104376 +/- 0.0105006

SVD, k=3



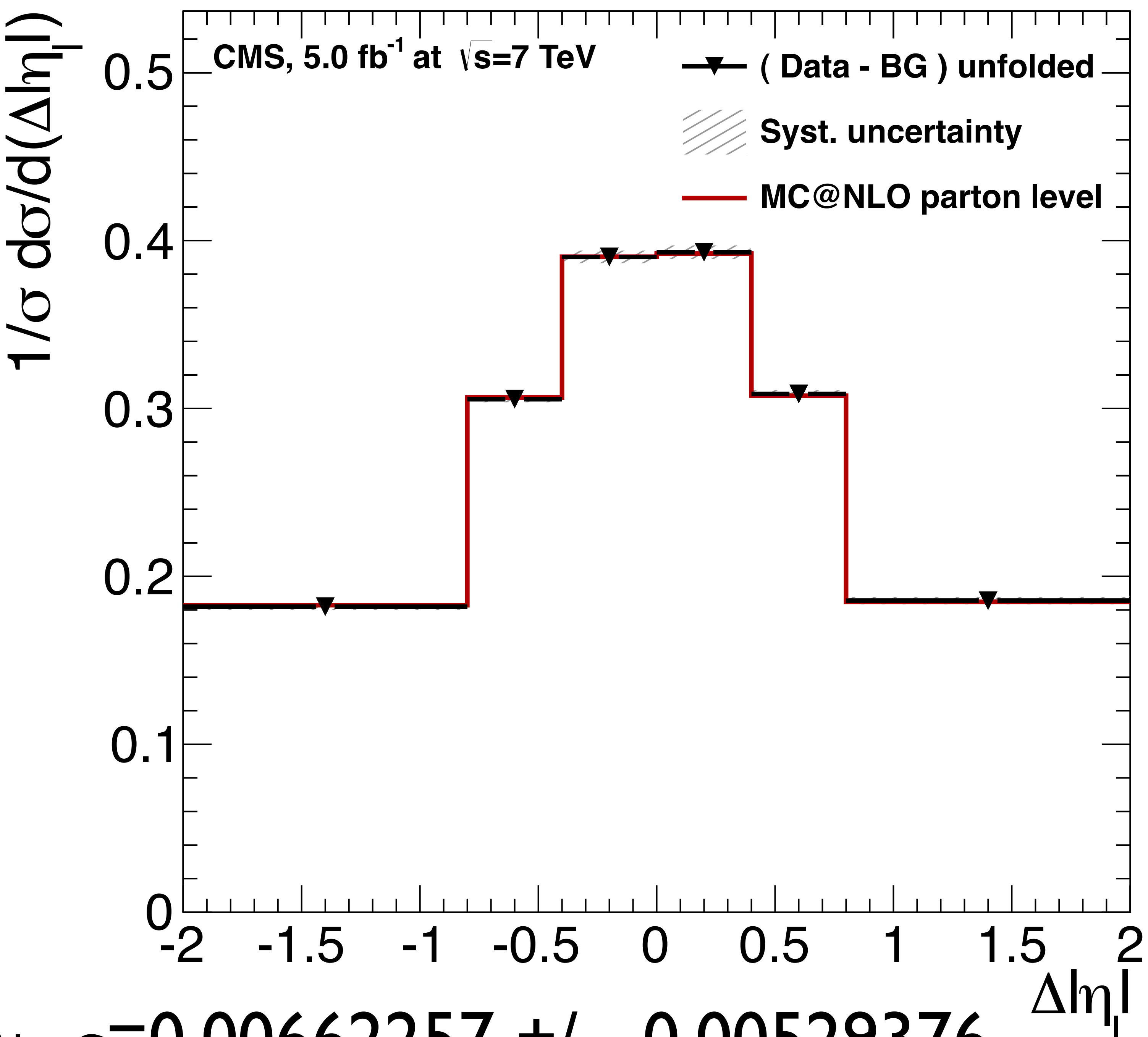
A_{lepC}=0.0092841 +/- 0.00984889

TUnfold, tau = 0.02



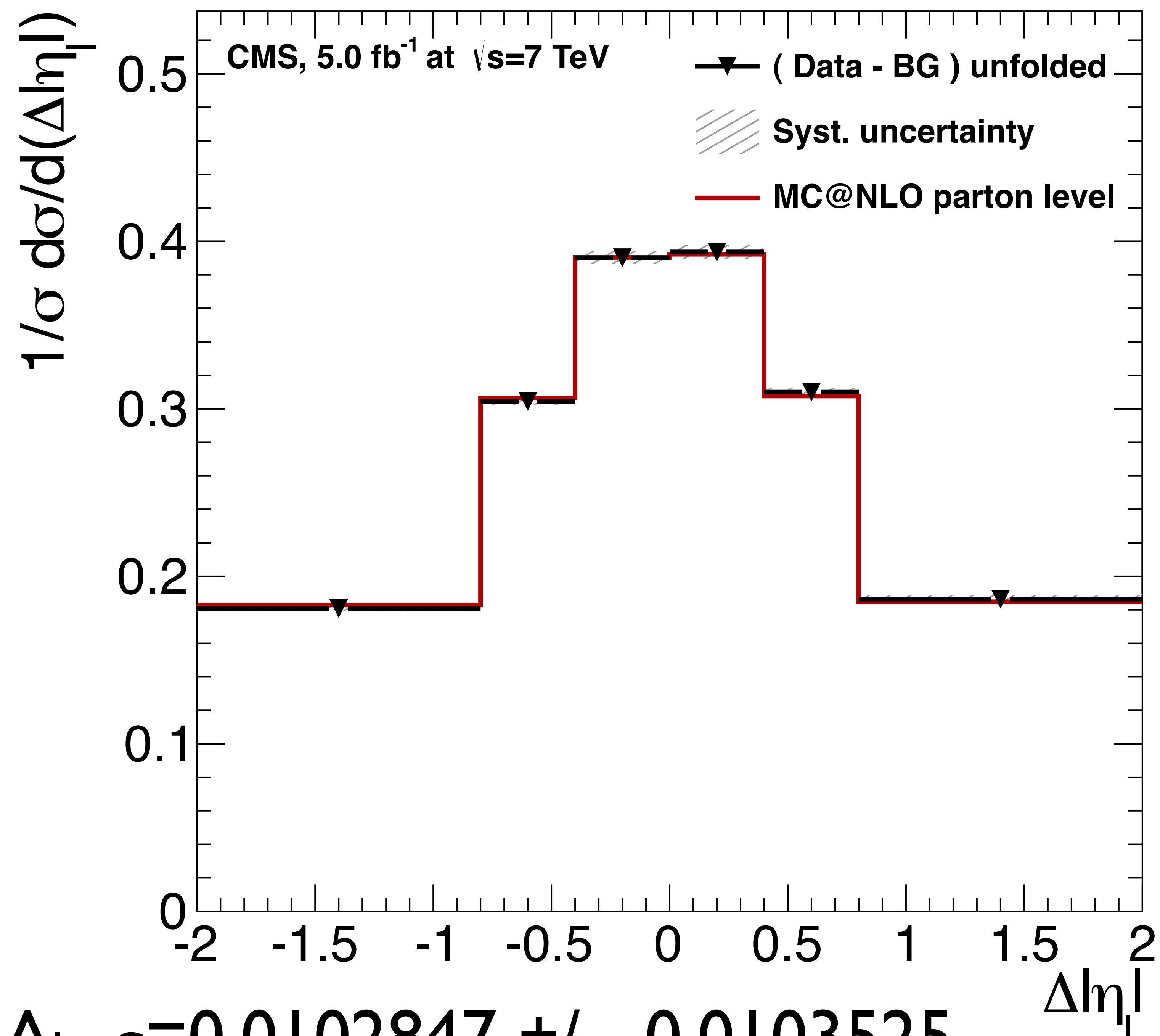
A_{lepC}=0.0104128 +/- 0.0103902

SVD, k=2



A_{lepC}=0.00662257 +/- 0.00529376

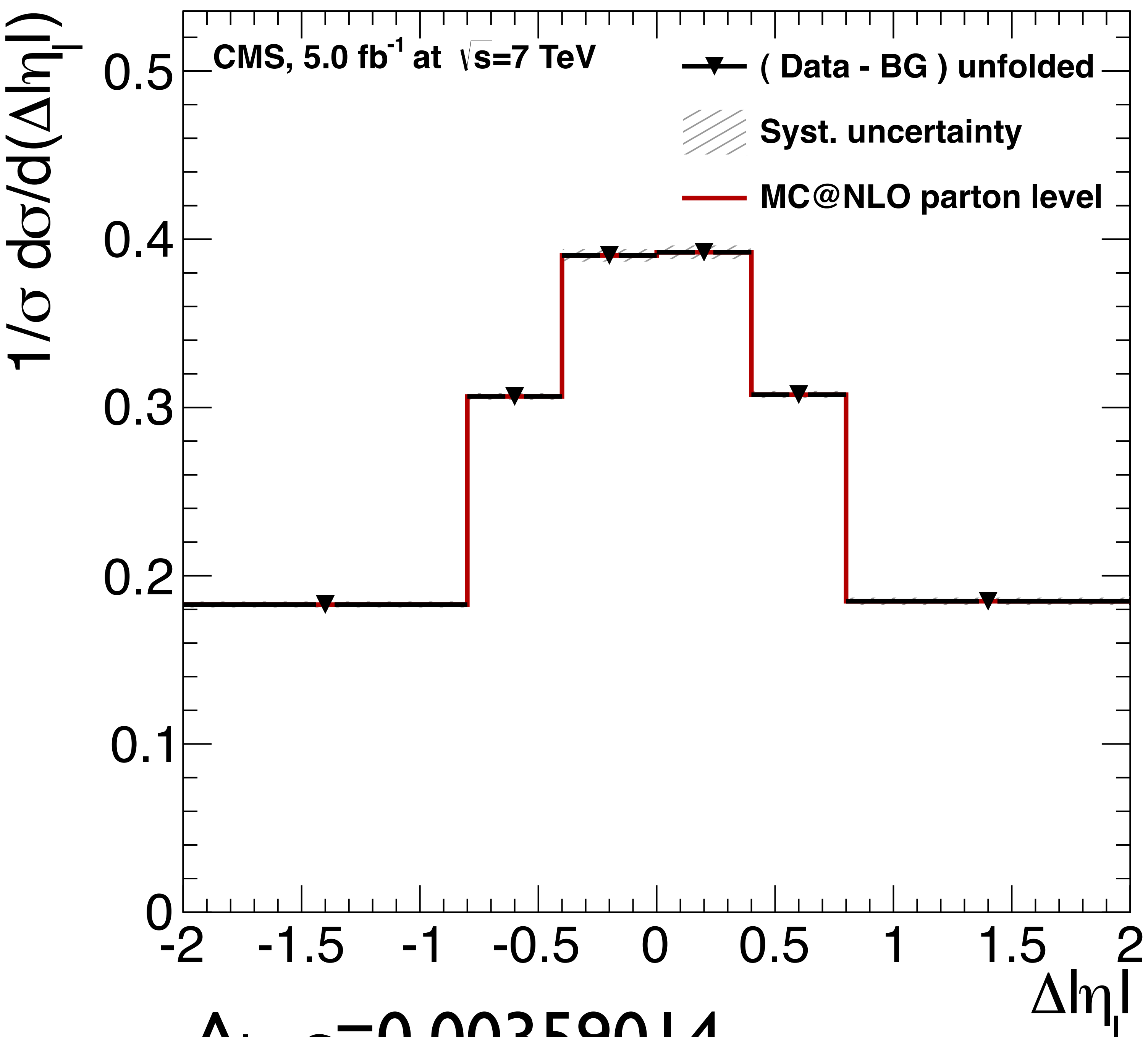
TUnfold, tau = 0.1

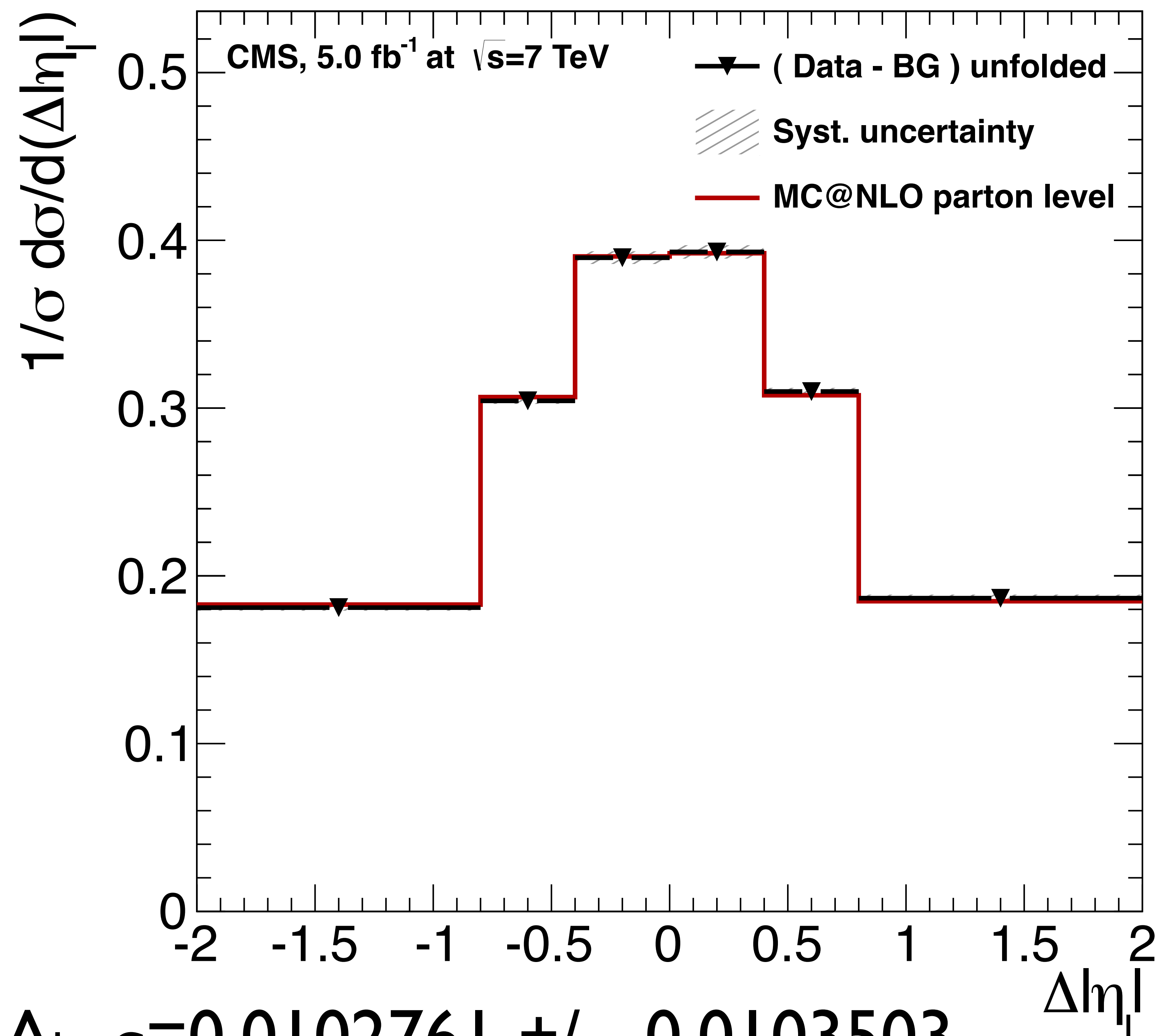


A_{lepC}=0.0102847 +/- 0.0103525

SVD, k=1

TUnfold, tau = 10



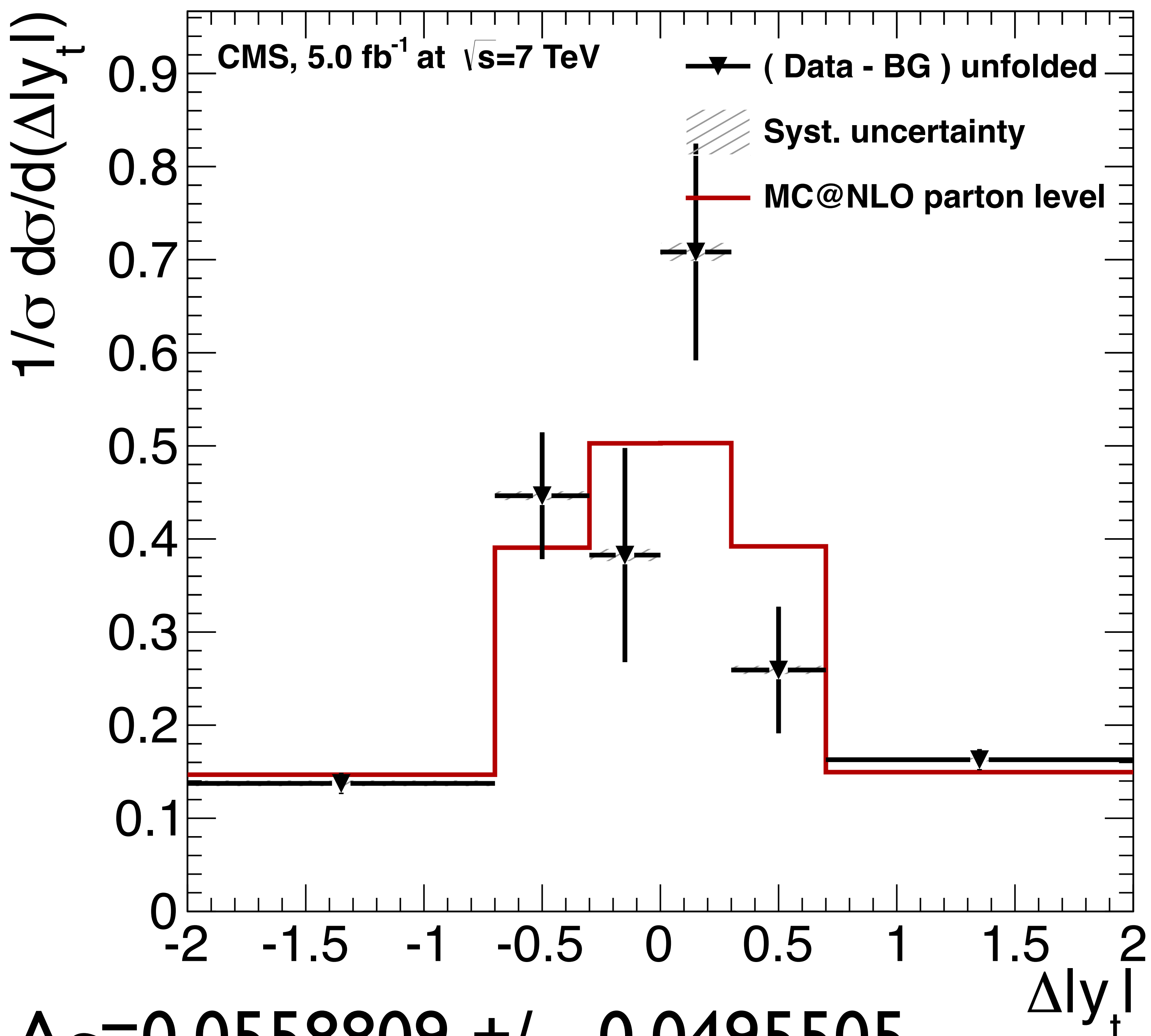


A_{lepC}=0.00359014

A_{lepC}=0.0102761 +/- 0.0103503

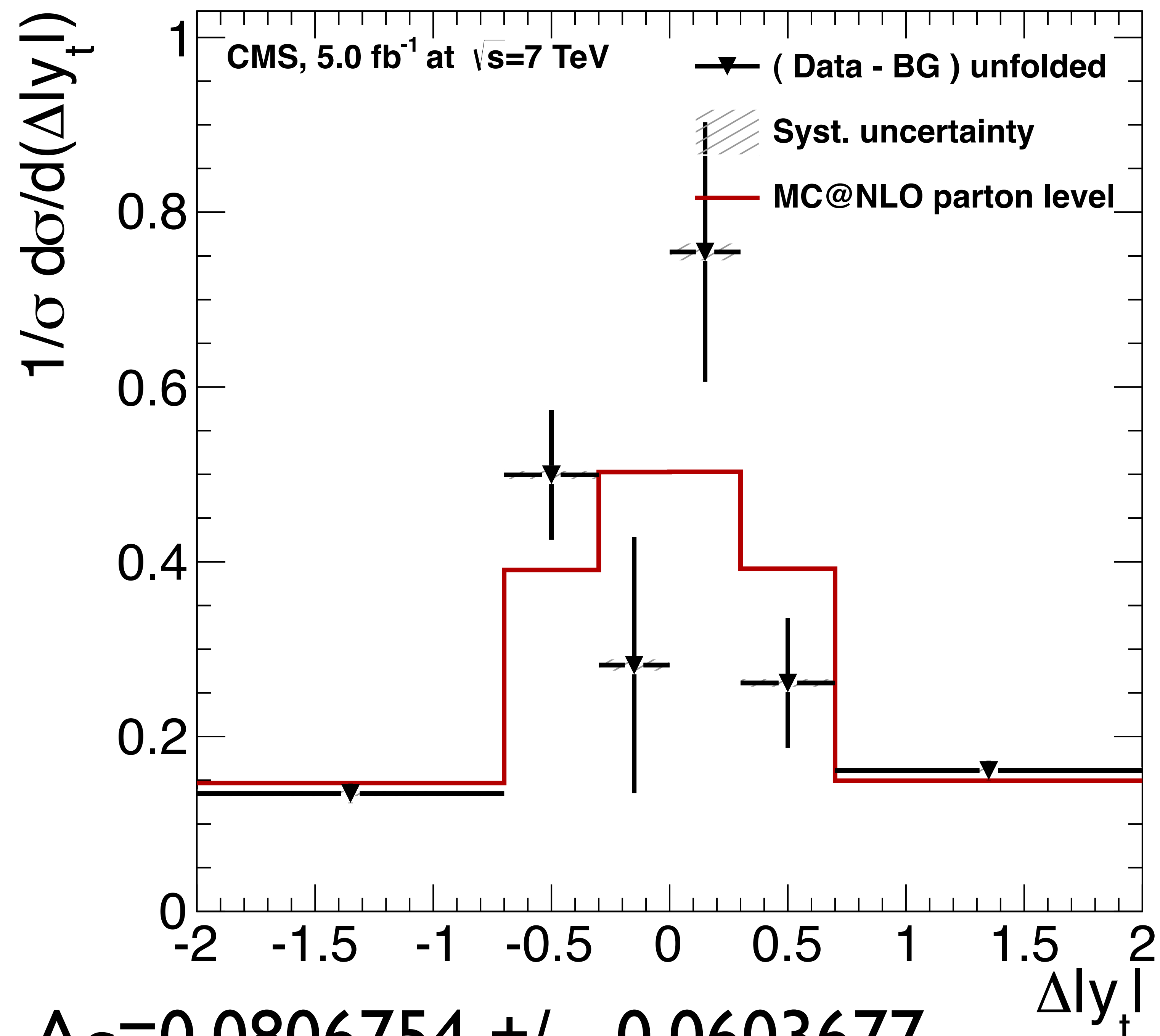
A_C results

SVD, k=6



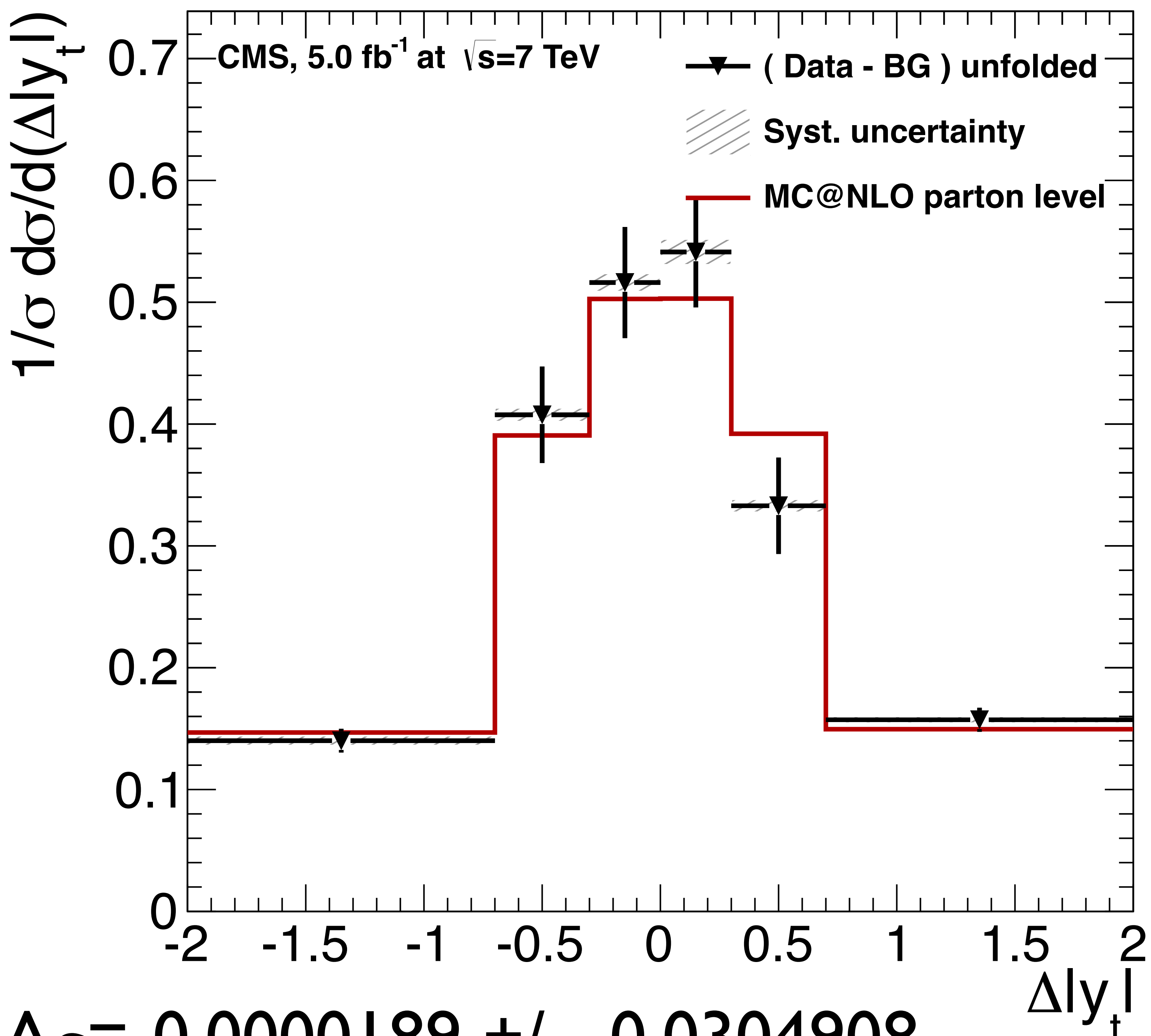
A_C=0.0558809 +/- 0.0495505

TUnfold, tau = 0.0003



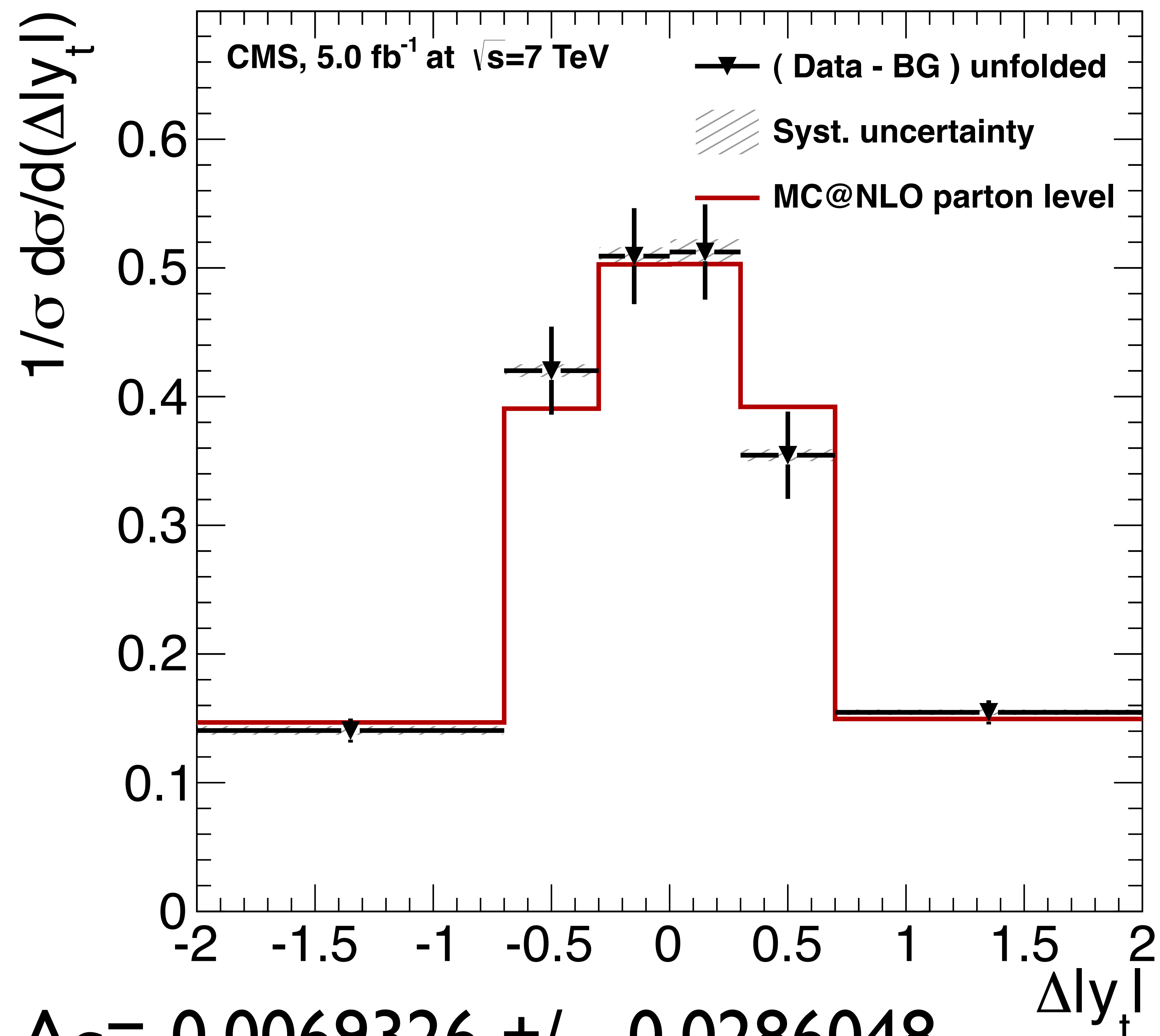
A_C=0.0806754 +/- 0.0603677

SVD, k=5



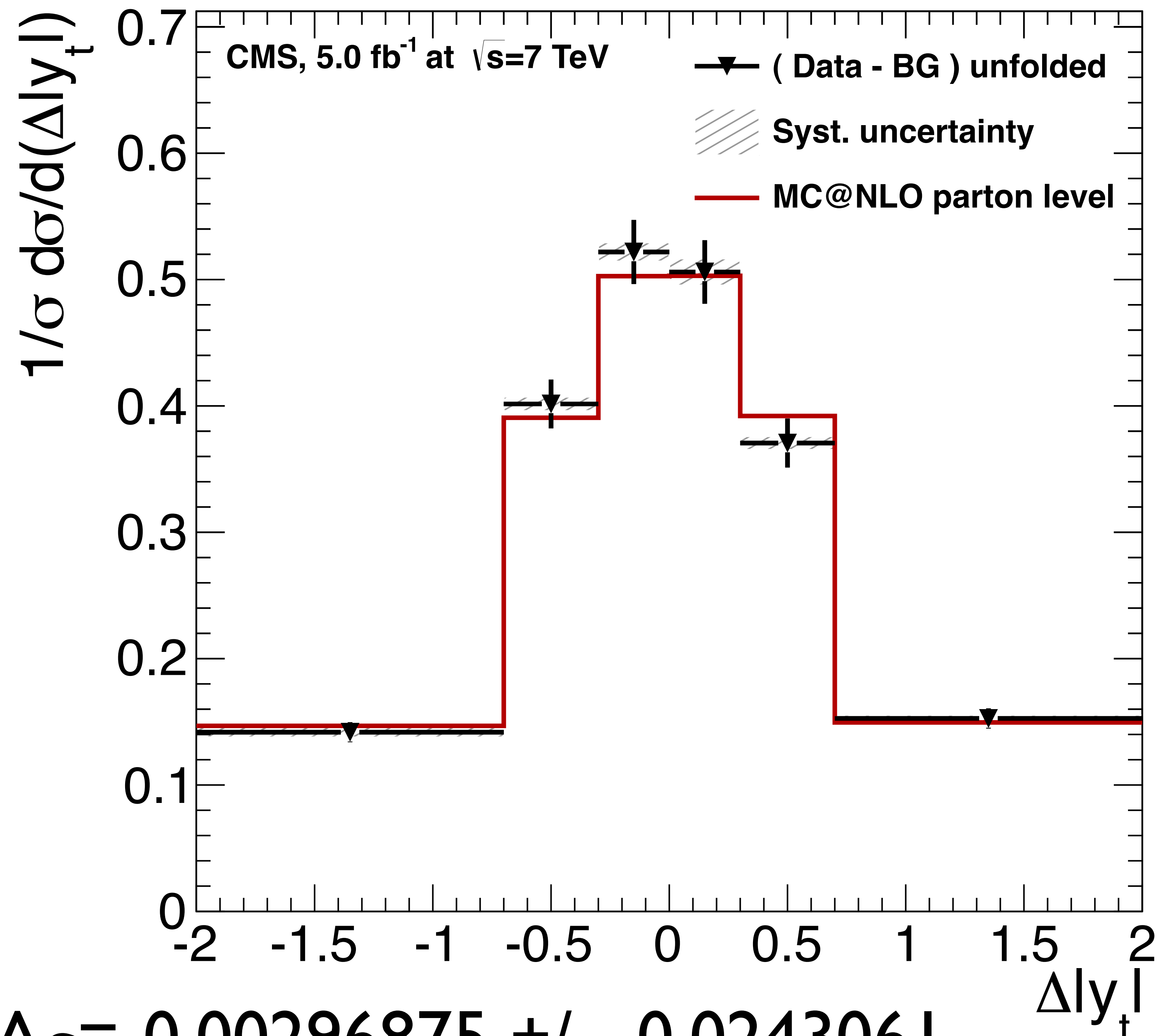
A_C=-0.0000189 +/- 0.0304908

TUnfold, tau = 0.002



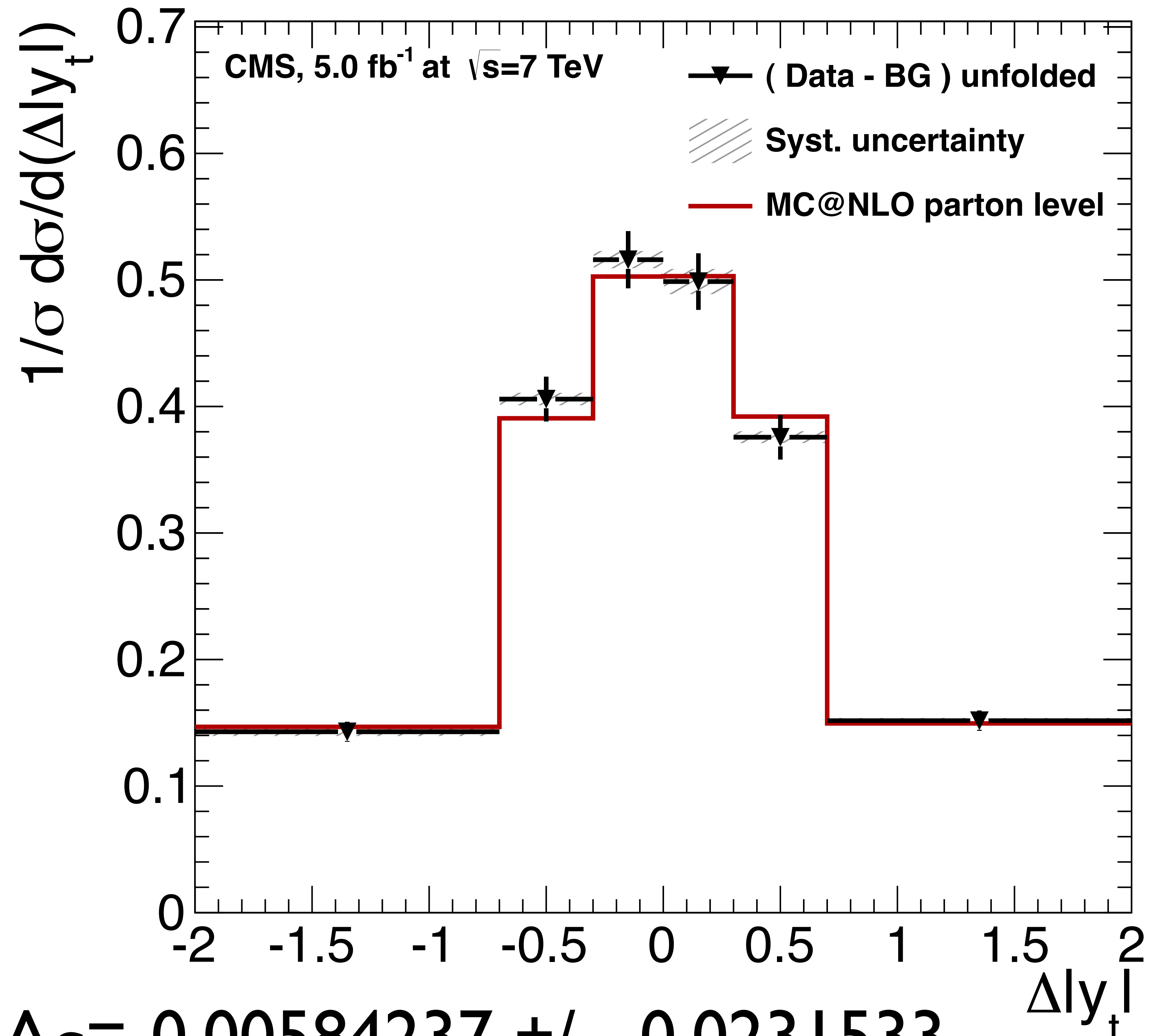
A_C=-0.0069326 +/- 0.0286048

SVD, k=4



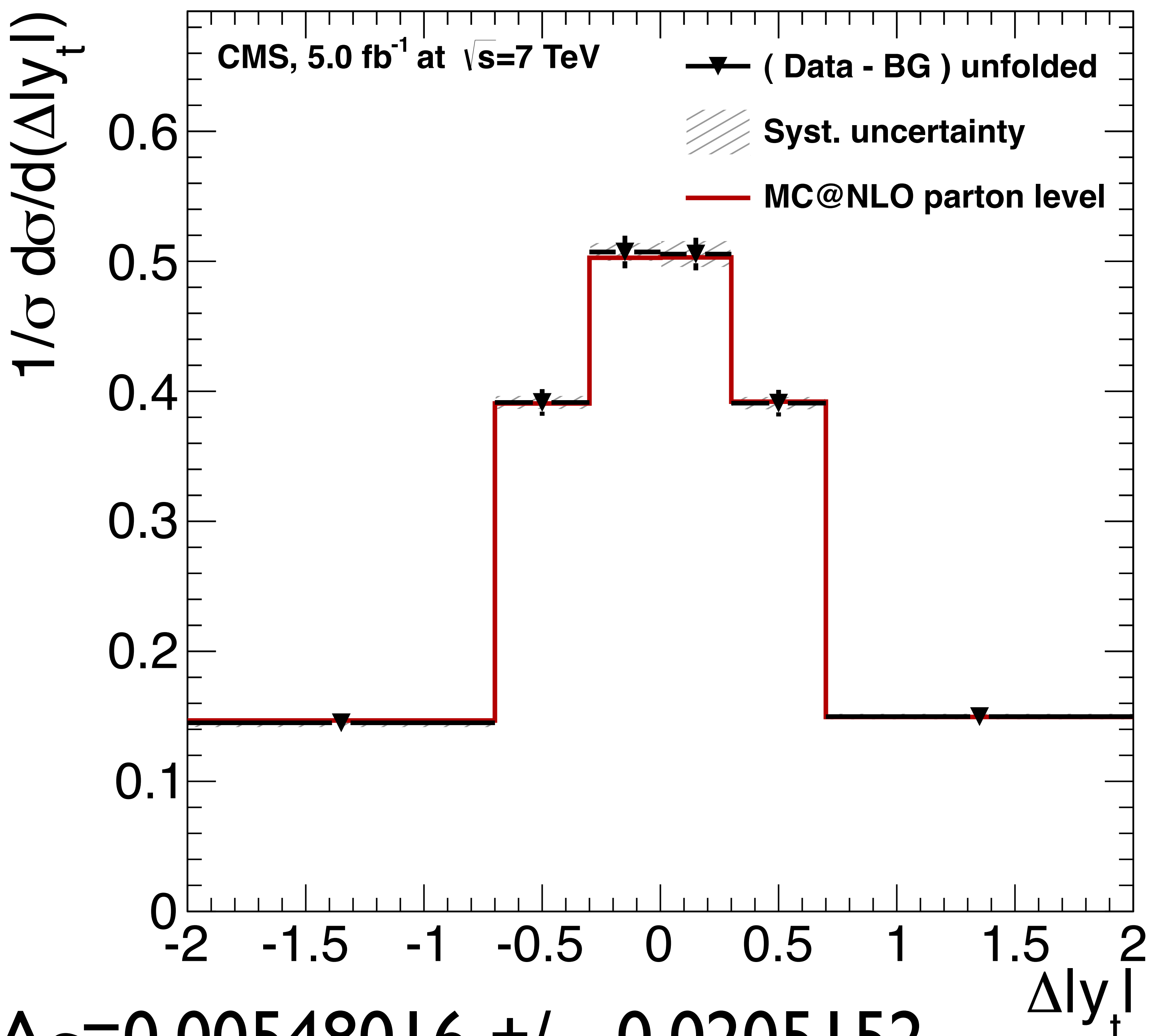
A_C=-0.00296875 +/- 0.0243061

TUnfold, tau = 0.005



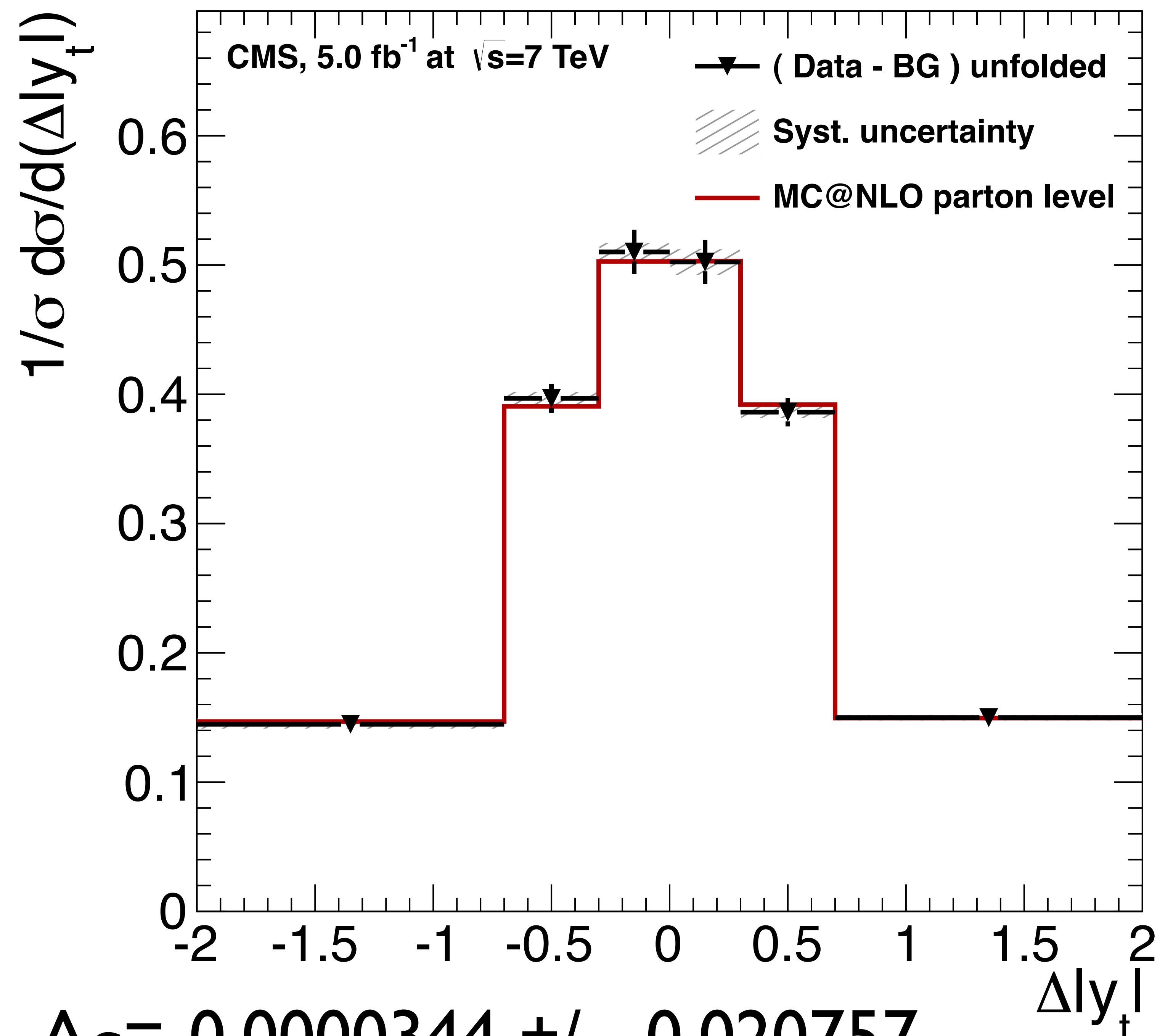
A_C=-0.00584237 +/- 0.0231533

SVD, k=3



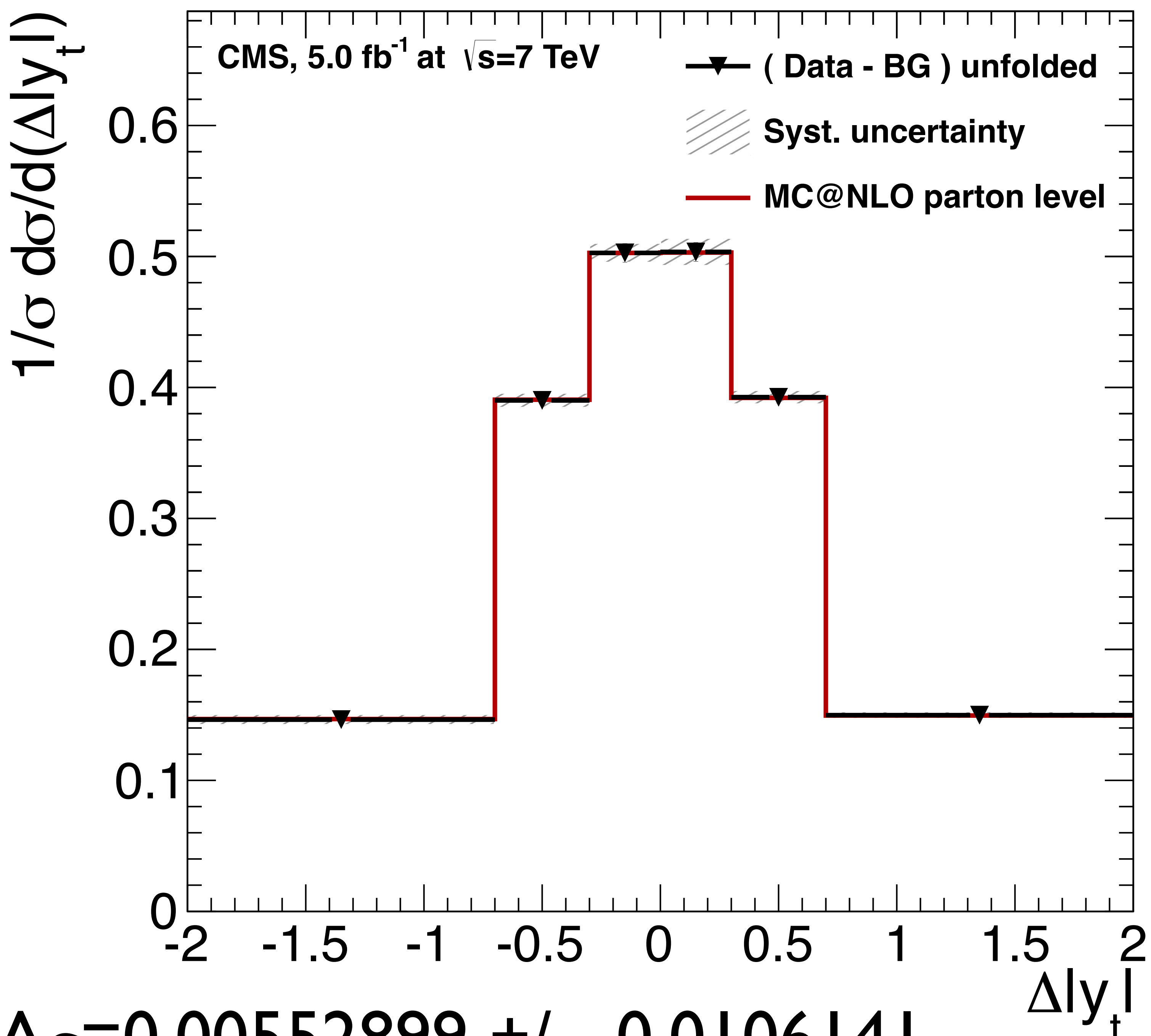
A_C=0.00548016 +/- 0.0205152

TUnfold, tau = 0.01



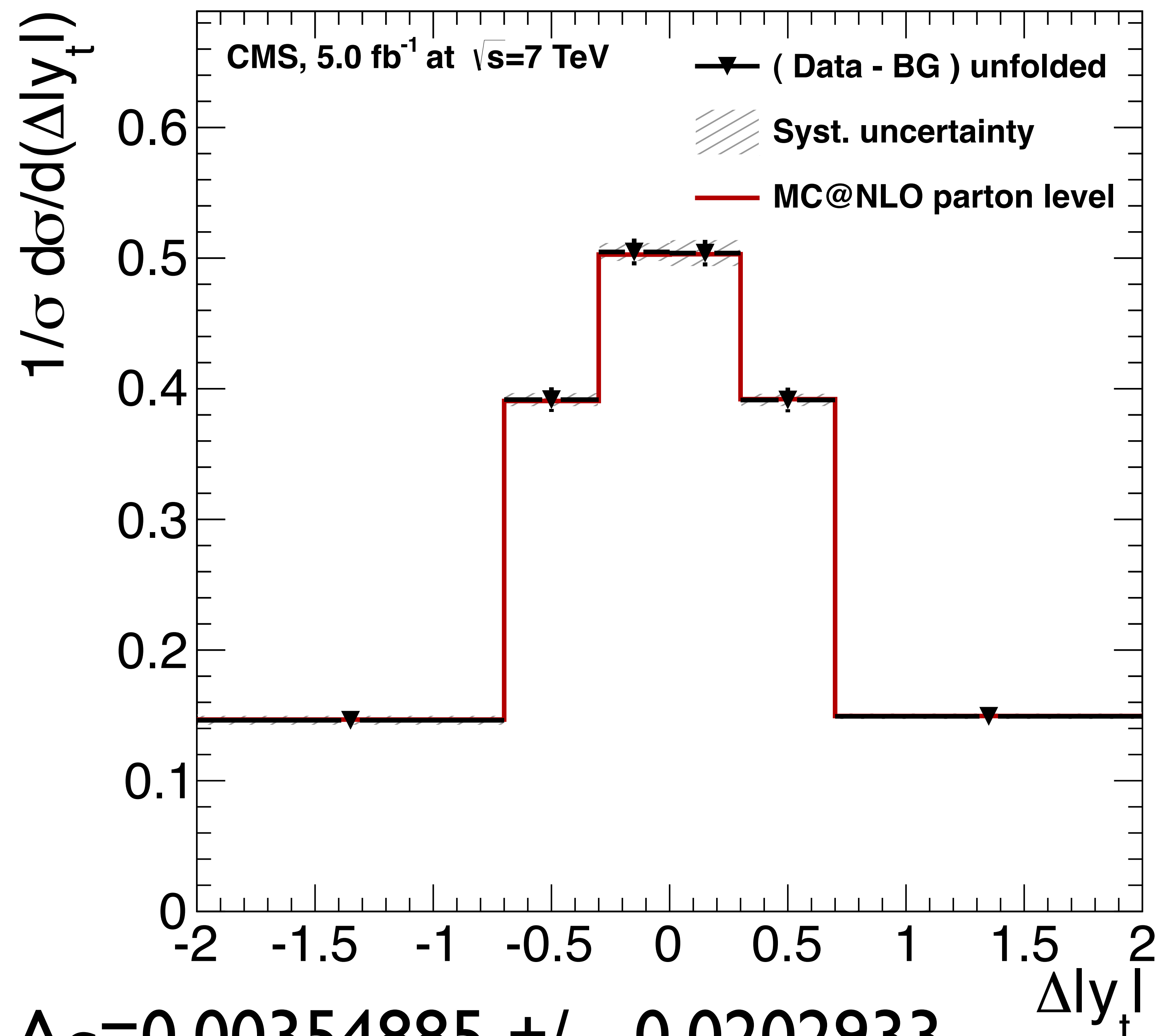
A_C=-0.0000344 +/- 0.020757

SVD, k=2



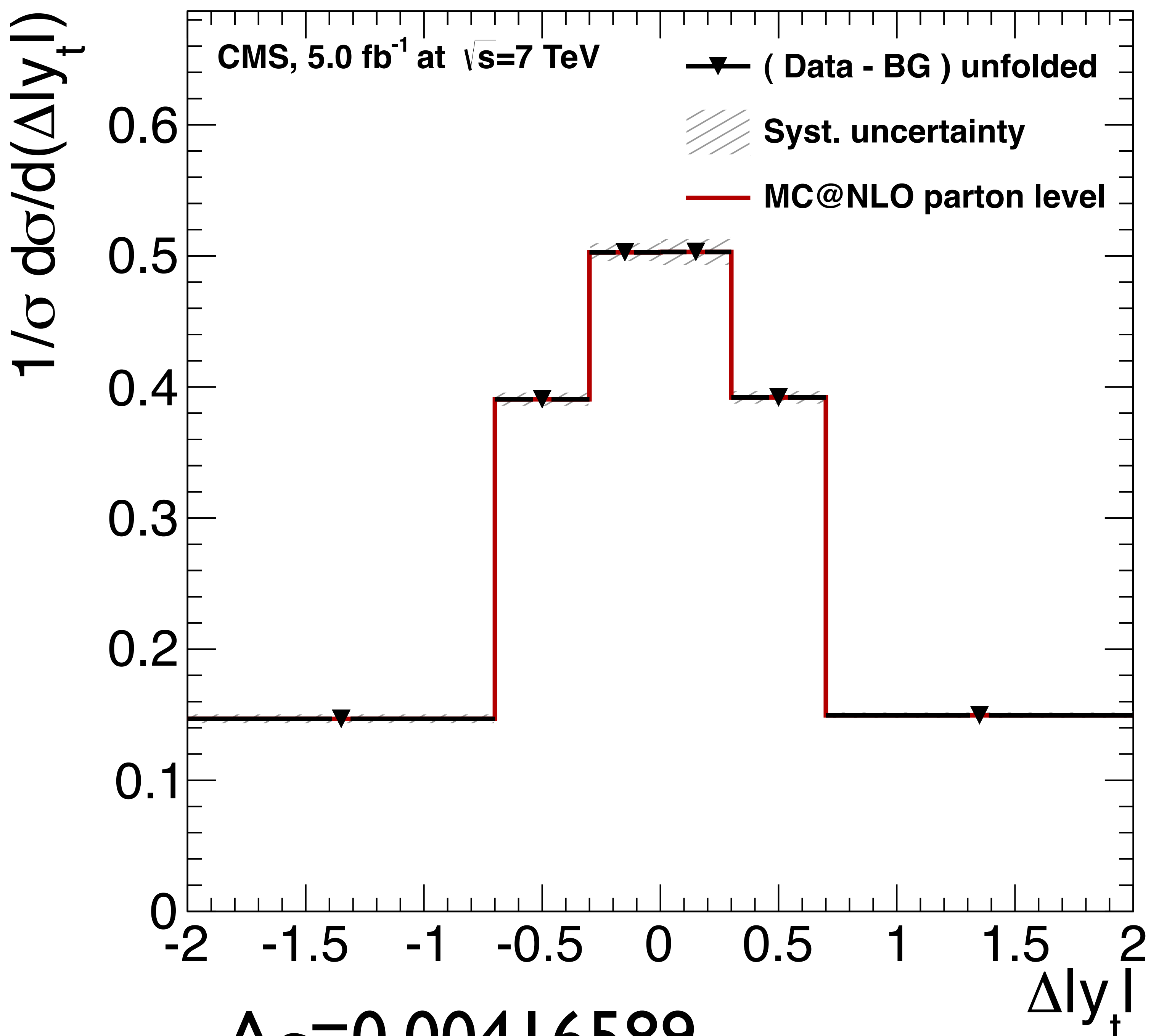
A_C=0.00552899 +/- 0.0106141

TUnfold, tau = 0.03



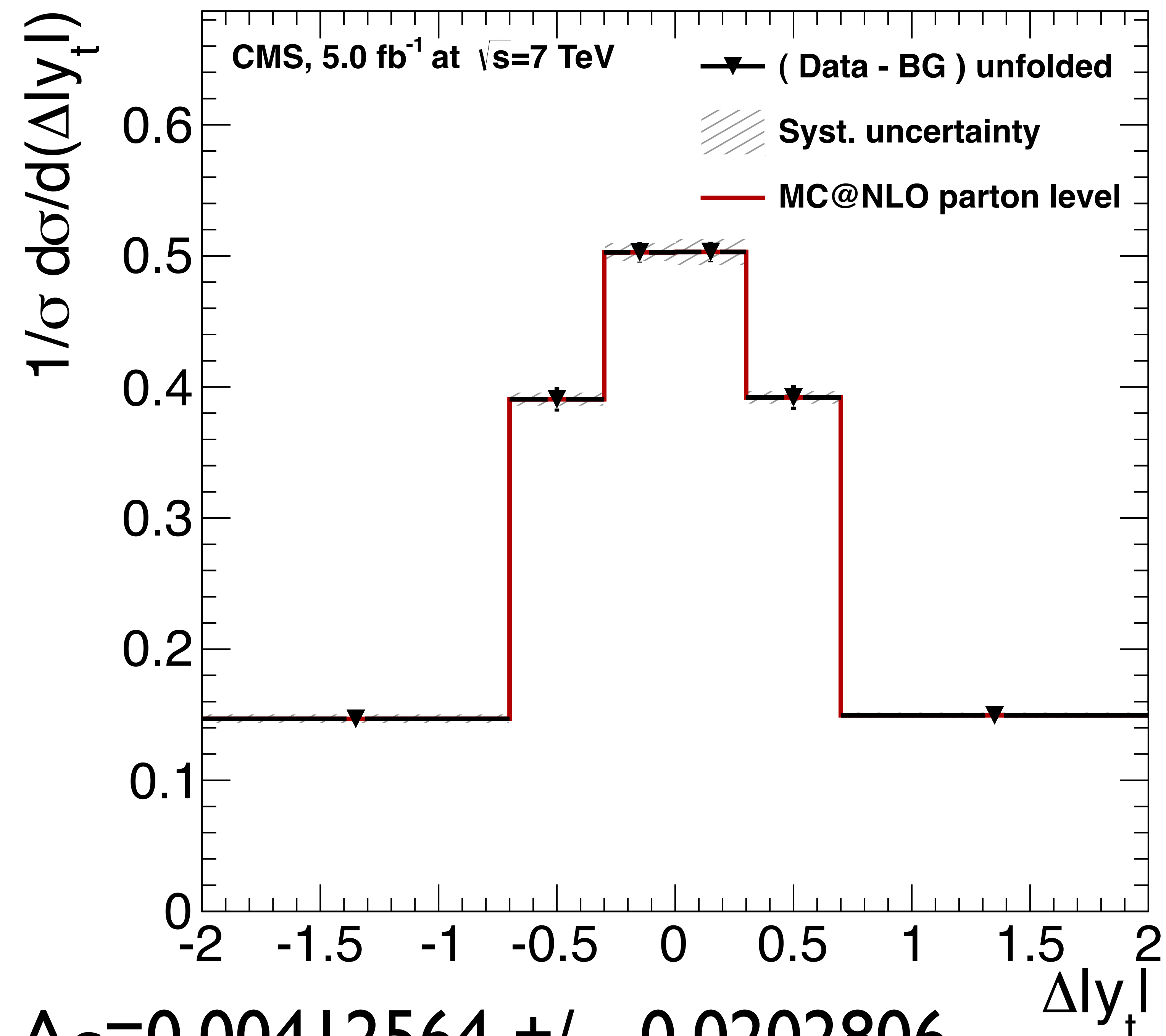
A_C=0.00354885 +/- 0.0202933

SVD, k=1



A_C=0.00416589

TUnfold, tau = 10

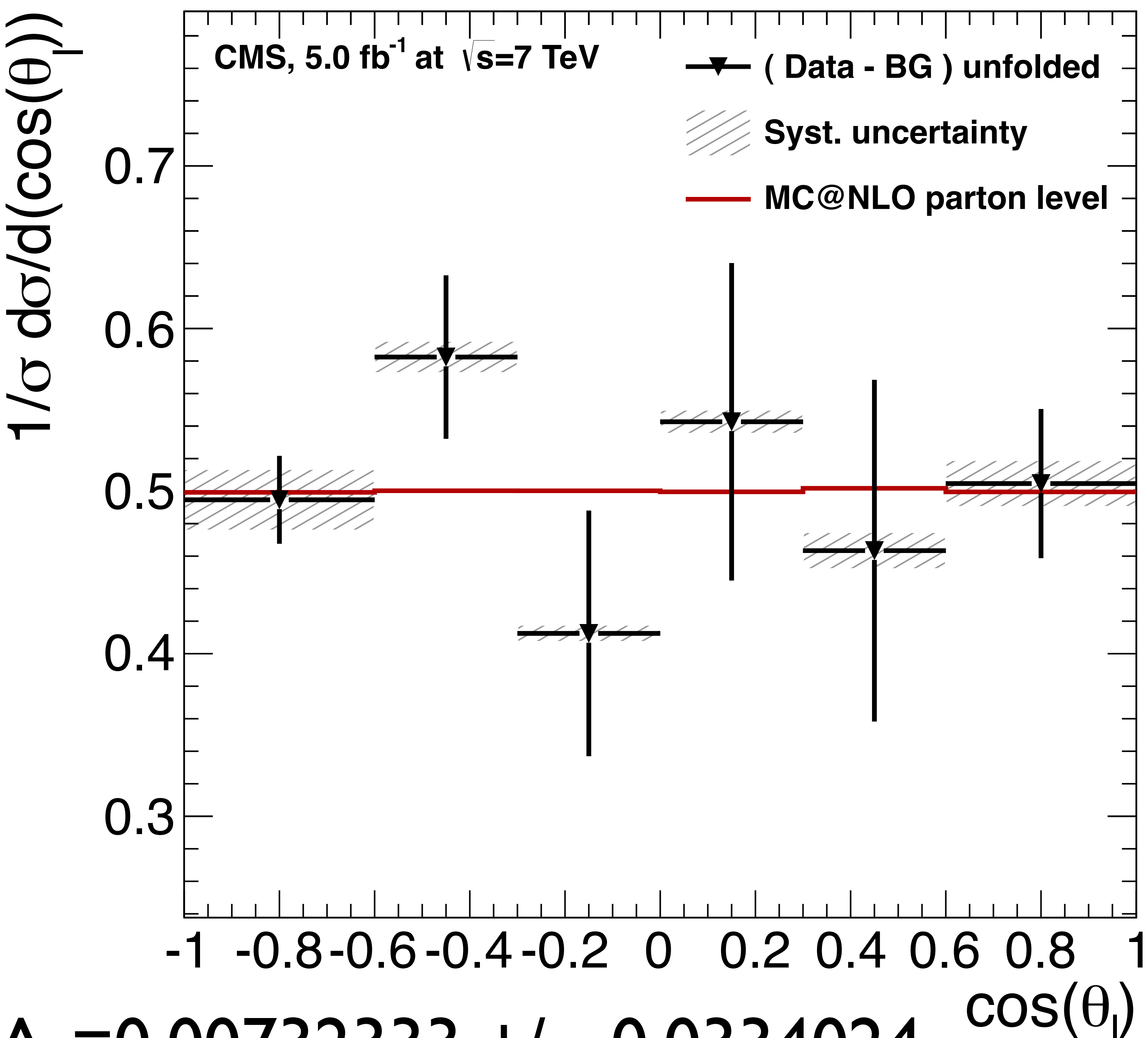


A_C=0.00412564 +/- 0.0202806

A_P results

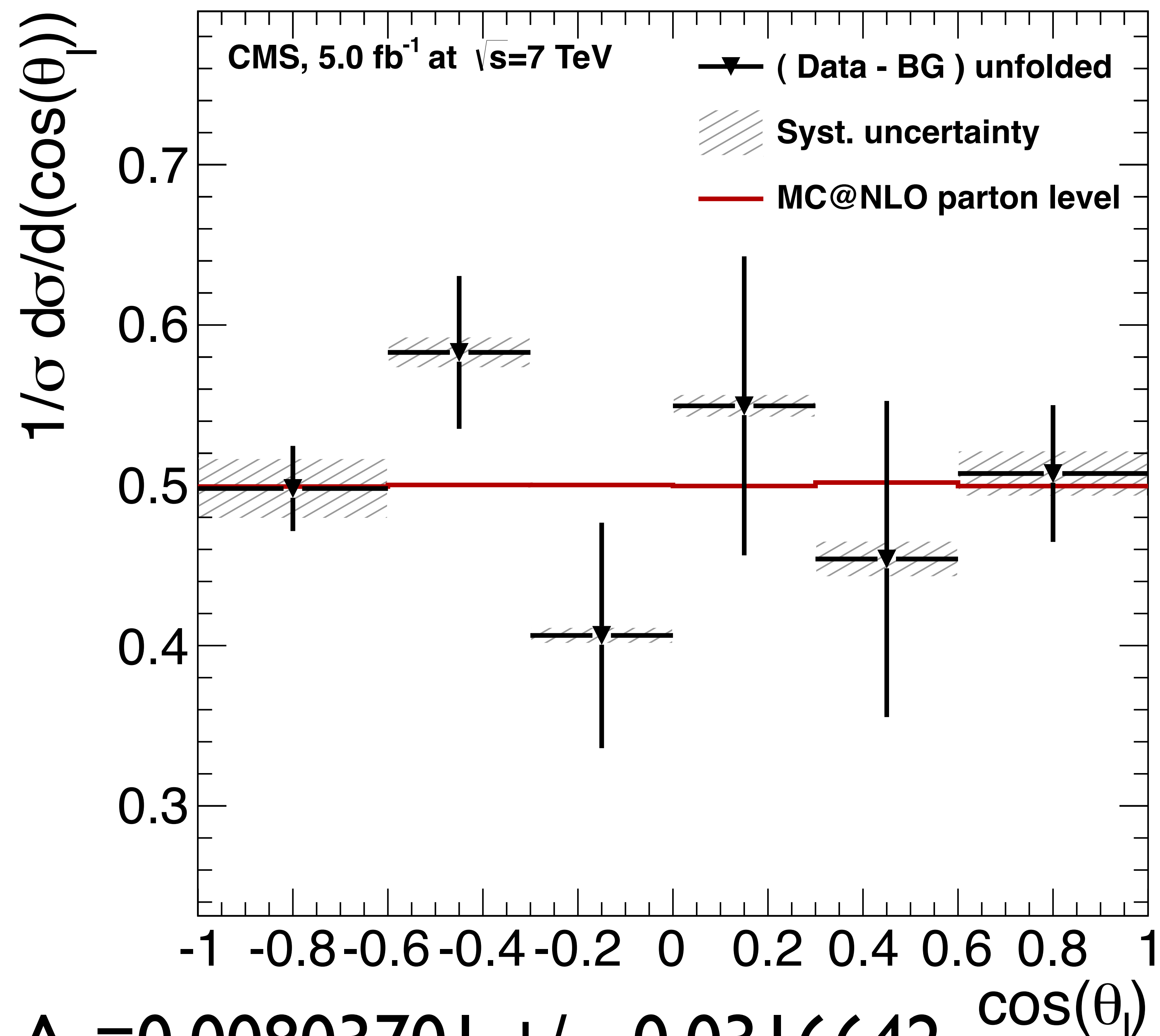
- Results are also similar for lepPlusCosTheta and lepMinusCosTheta individually

SVD, k=6



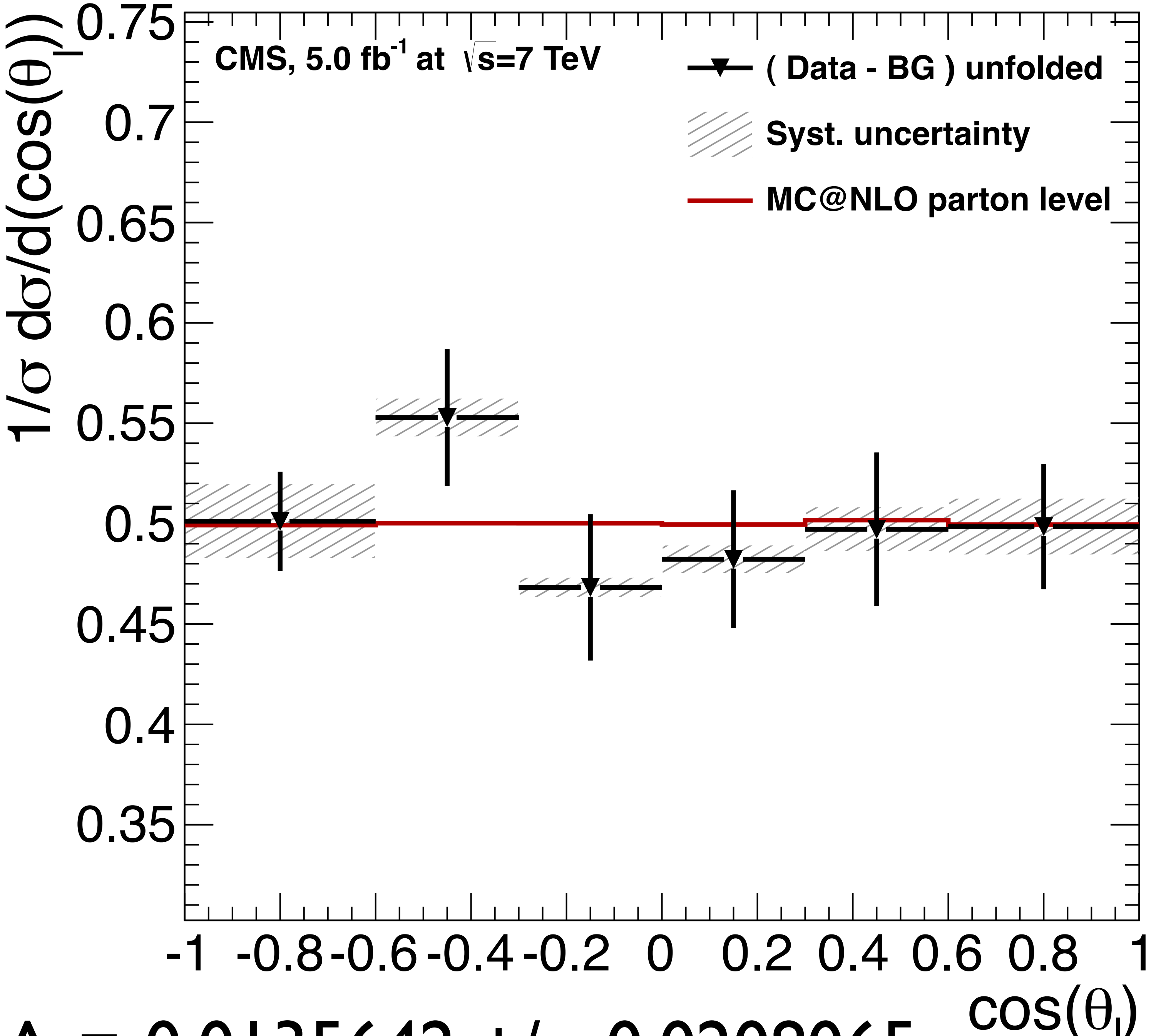
A_P=0.00732333 +/- 0.0334024

TUnfold, tau = 0.0003



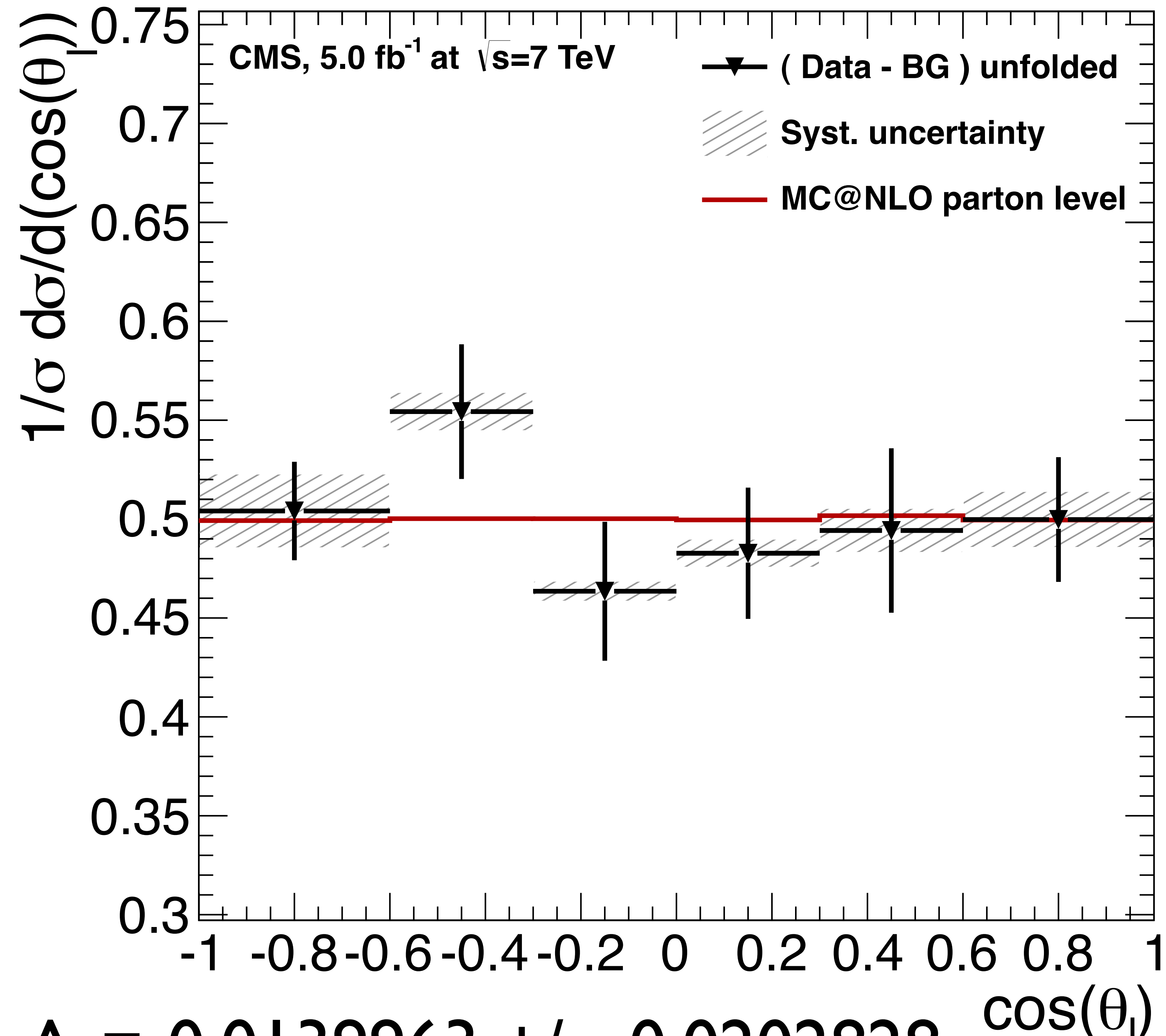
A_P=0.00803701 +/- 0.0316642

SVD, k=5



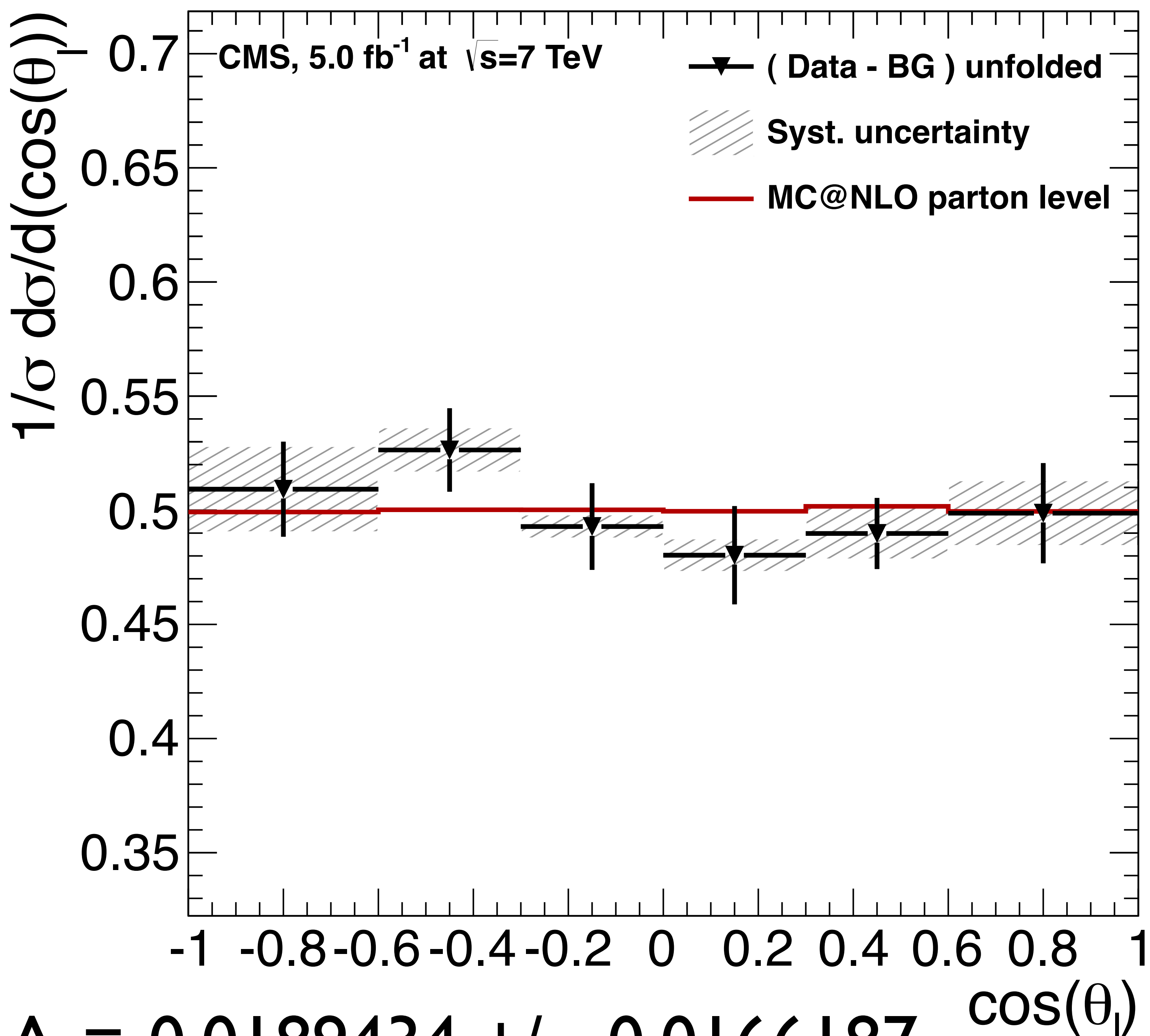
A_P=-0.0135642 +/- 0.0208065

TUnfold, tau = 0.001



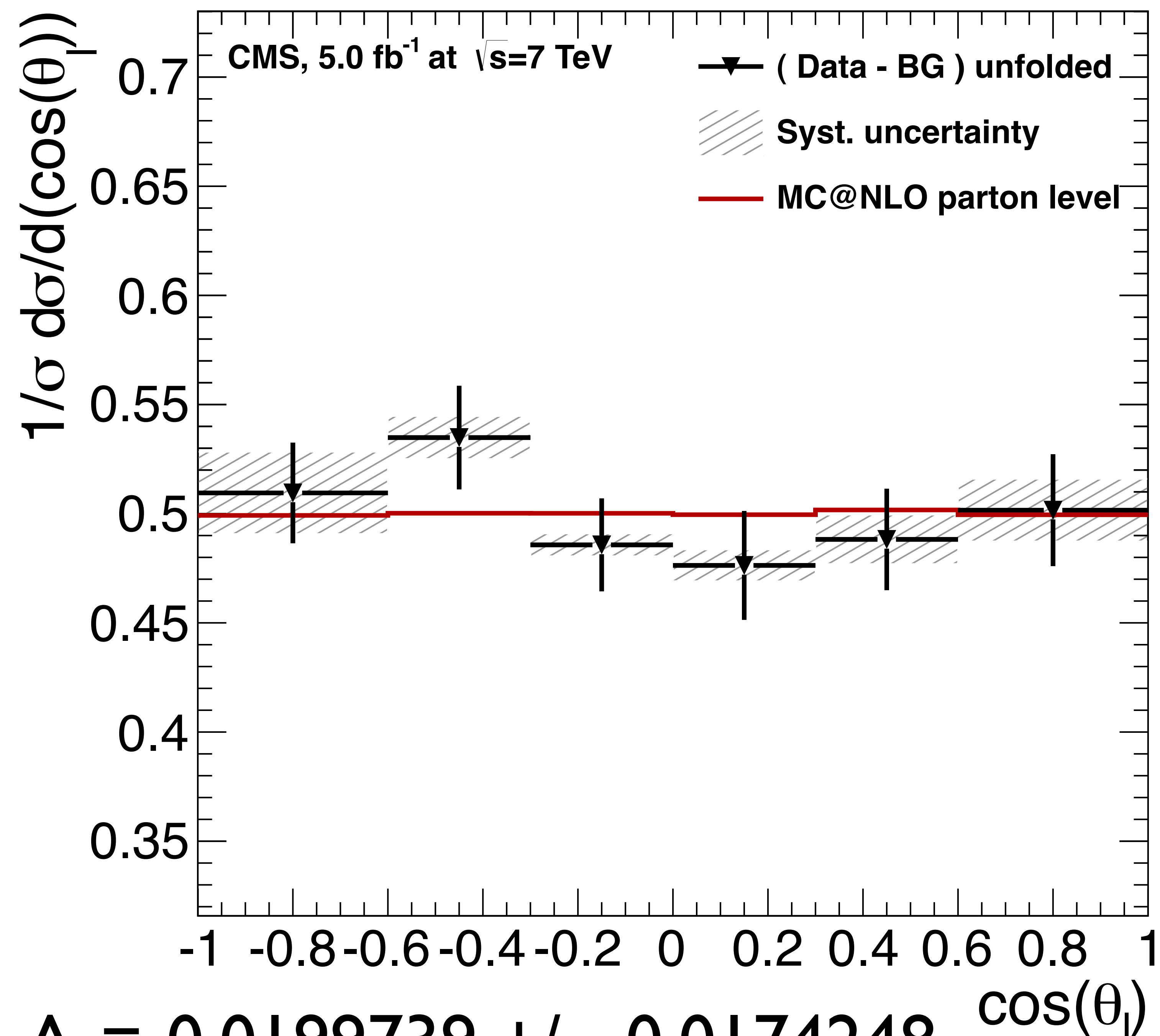
A_P=-0.0139963 +/- 0.0202828

SVD, k=4



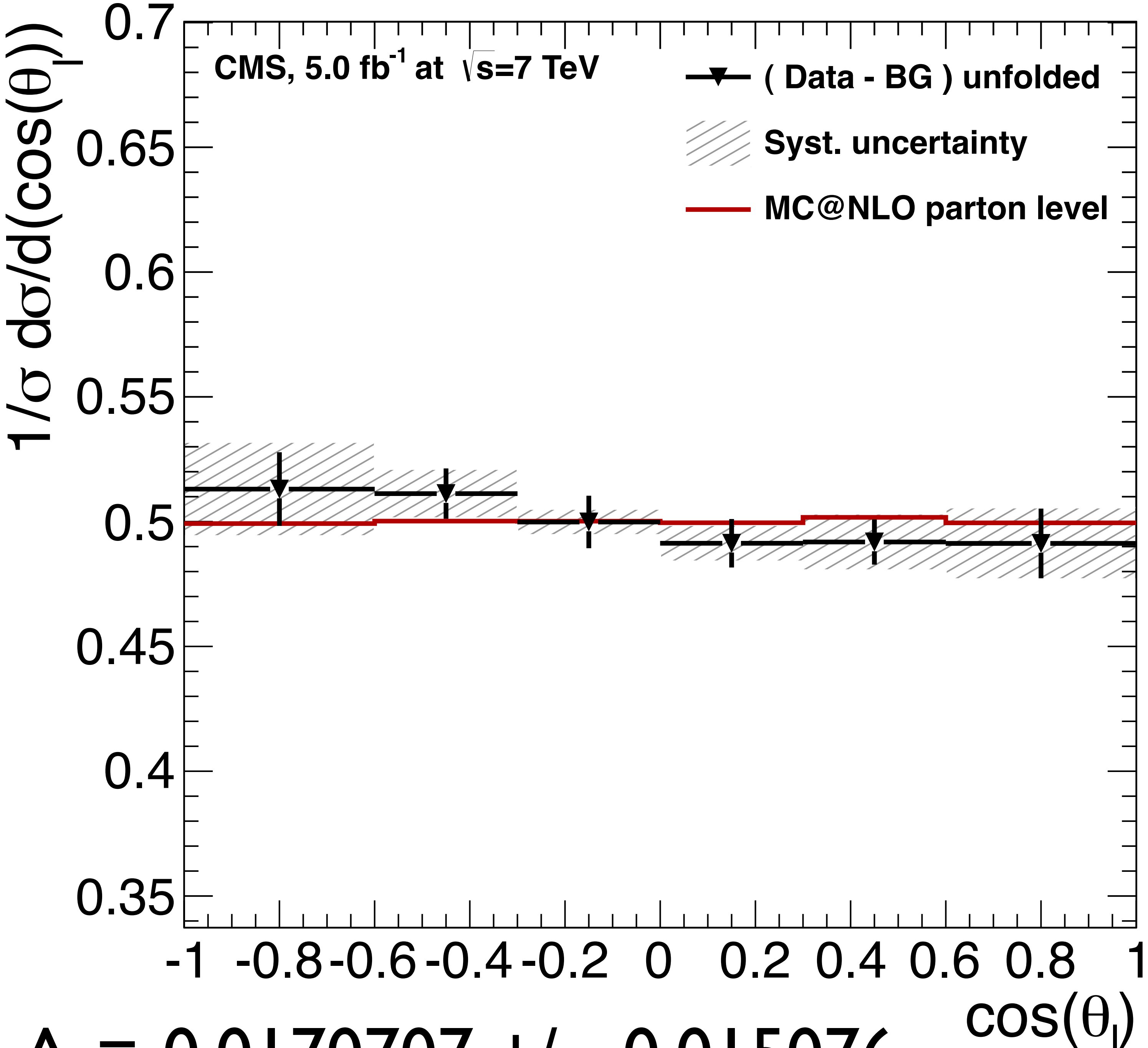
A_P=-0.0189434 +/- 0.0166187

TUnfold, tau = 0.002



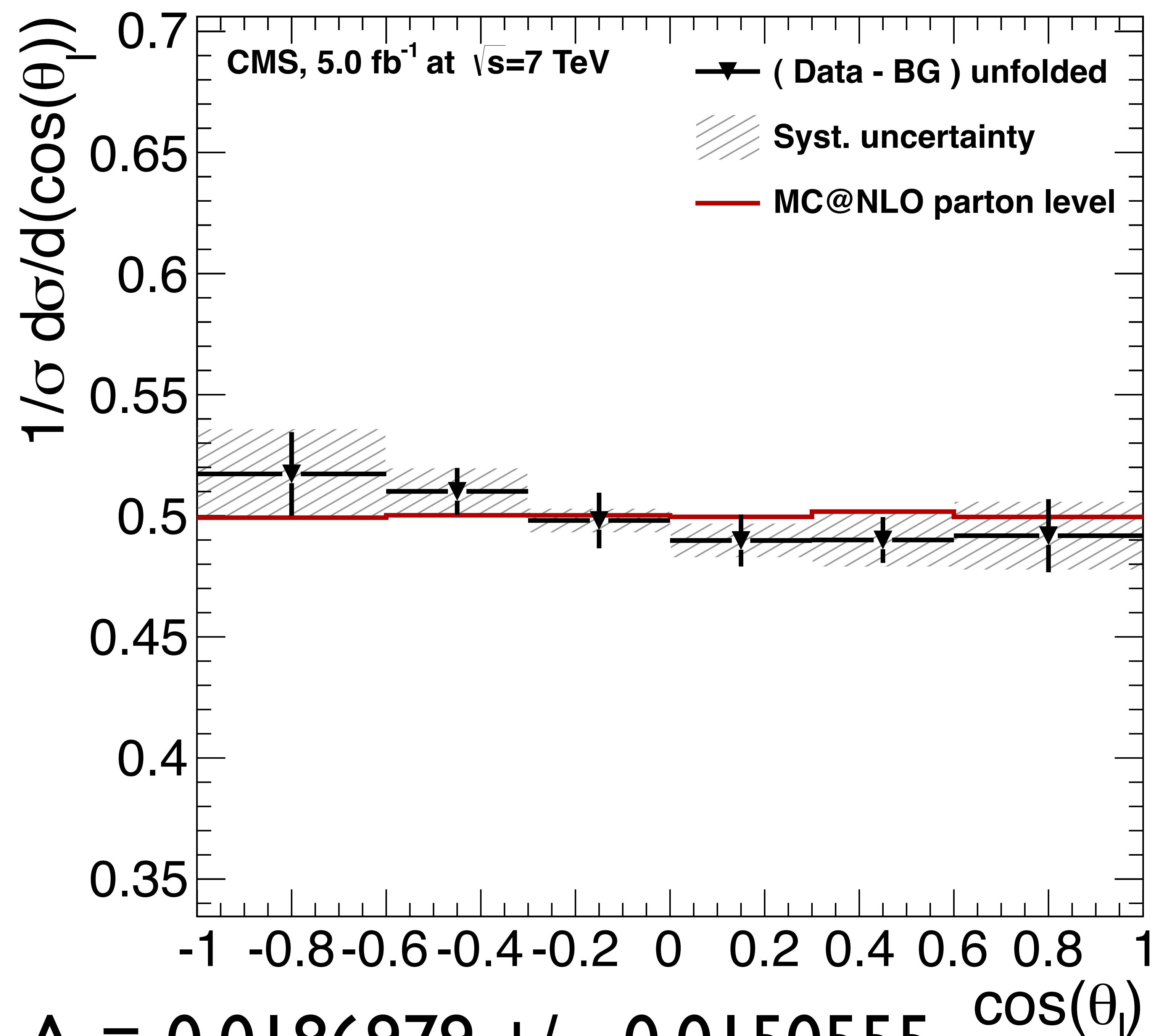
A_P=-0.0199739 +/- 0.0174248

SVD, k=3



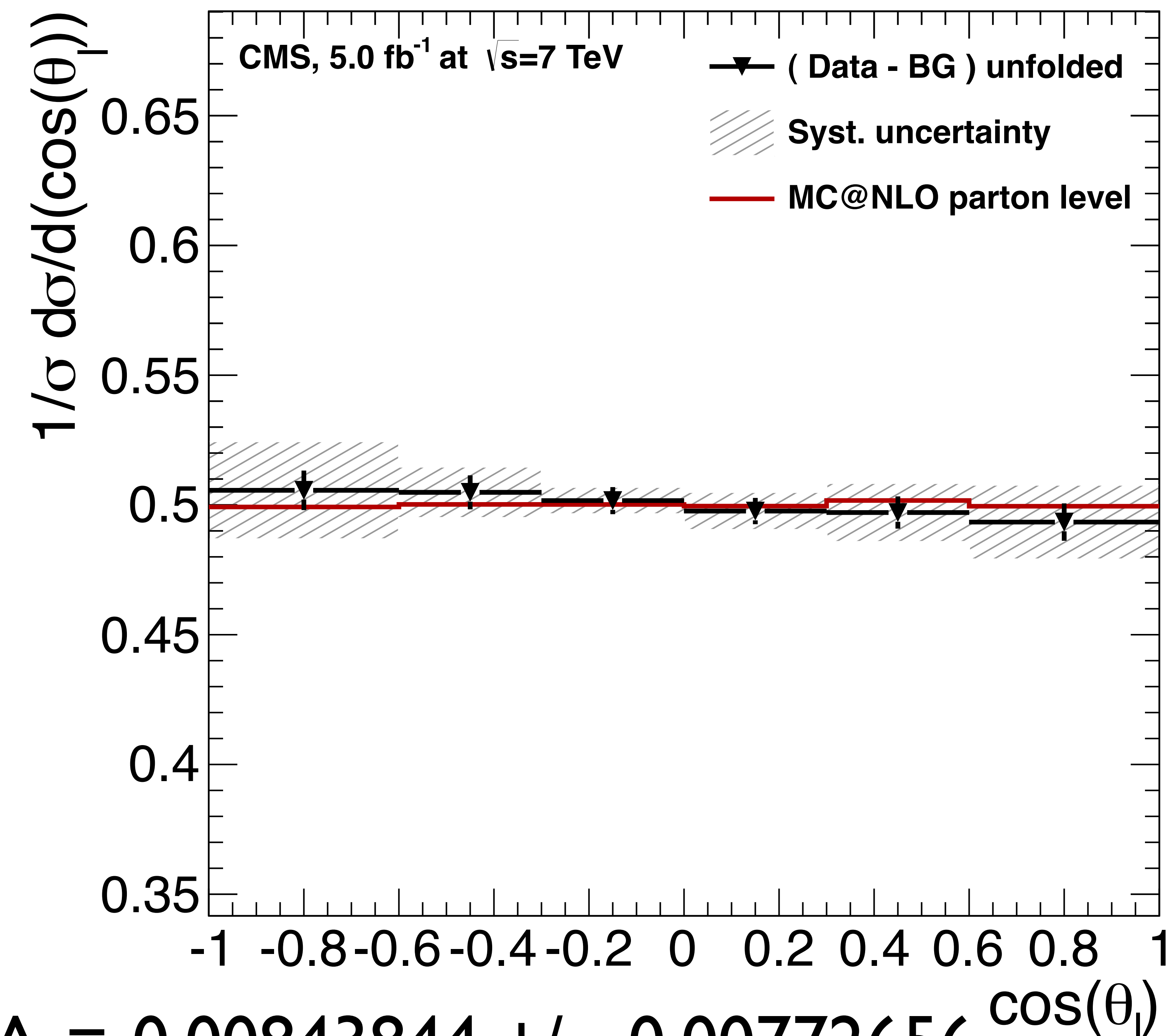
A_P=-0.0170707 +/- 0.015076

TUnfold, tau = 0.01



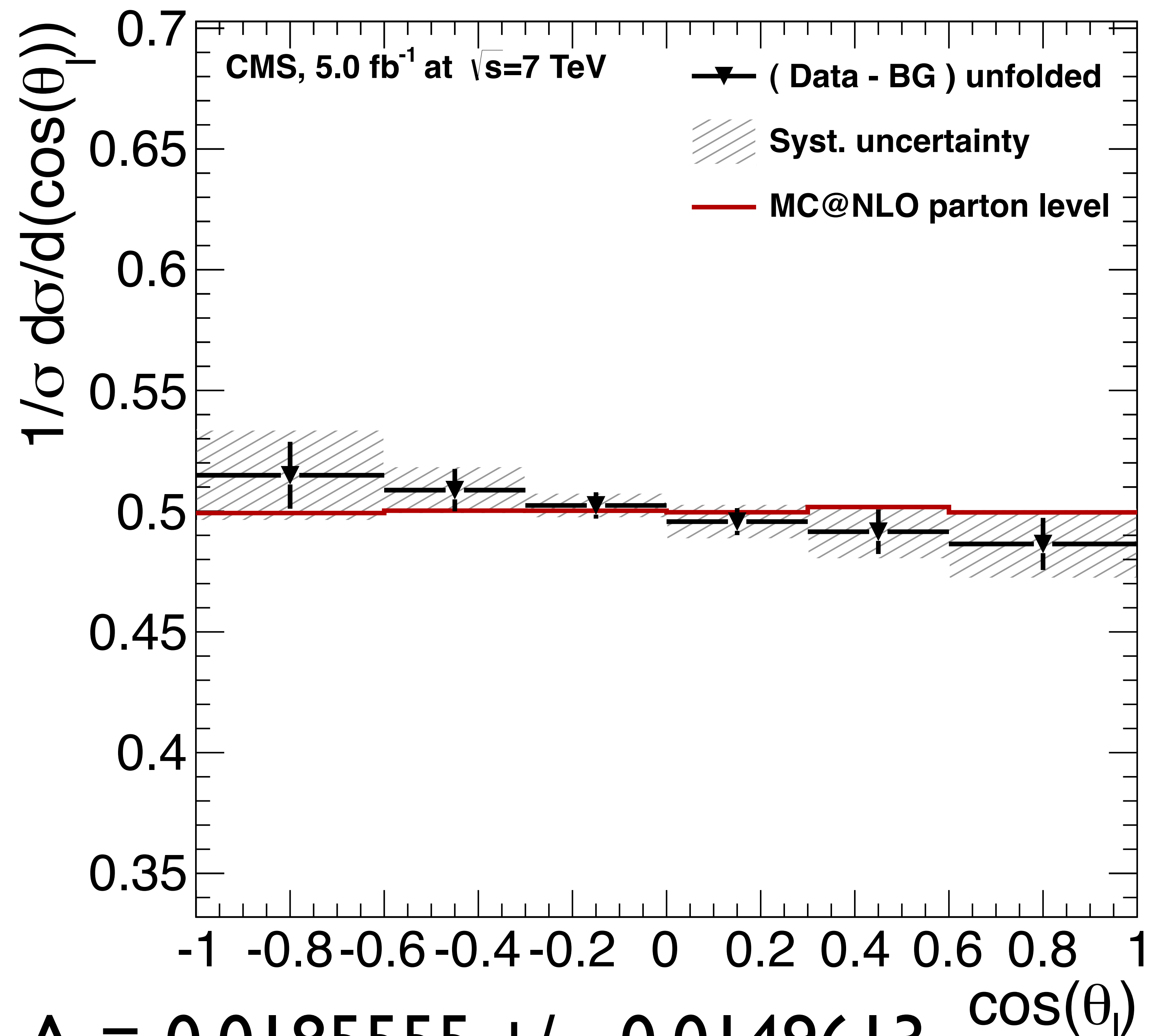
A_P=-0.0186979 +/- 0.0150555

SVD, k=2



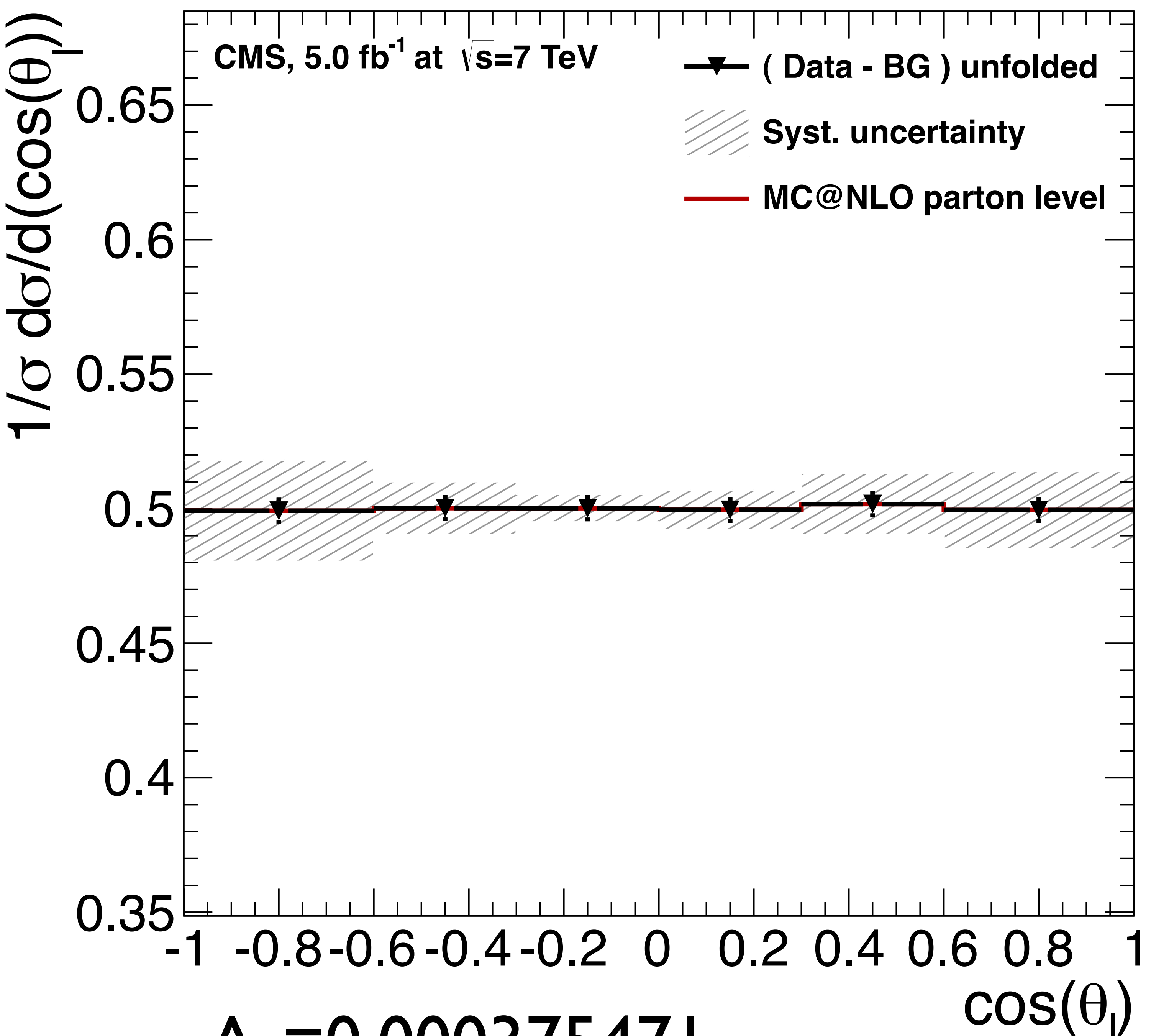
A_P=-0.00843844 +/- 0.00772656

TUnfold, tau = 0.1



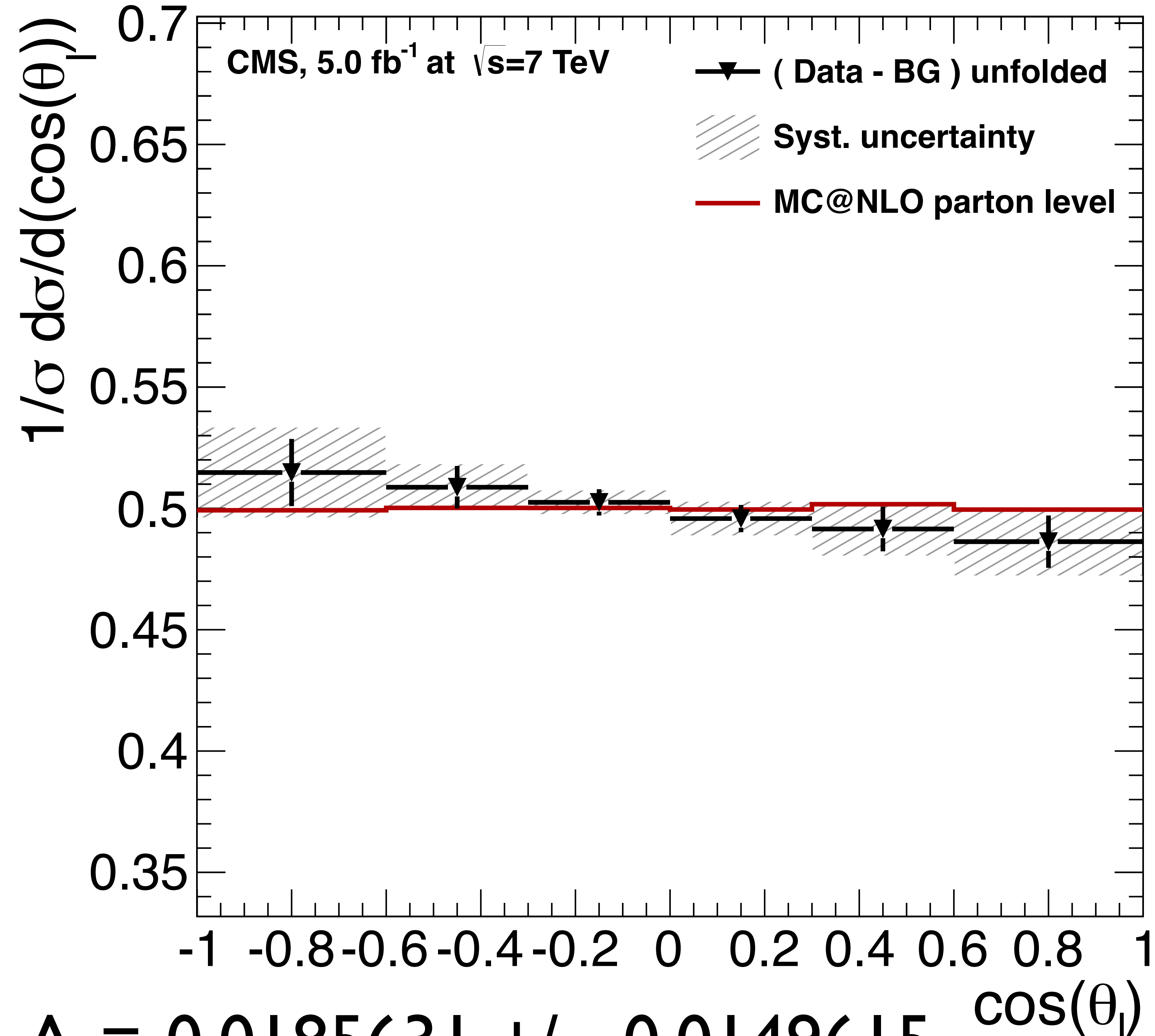
A_P=-0.0185555 +/- 0.0149613

SVD, k=1



A_P=0.00037547I

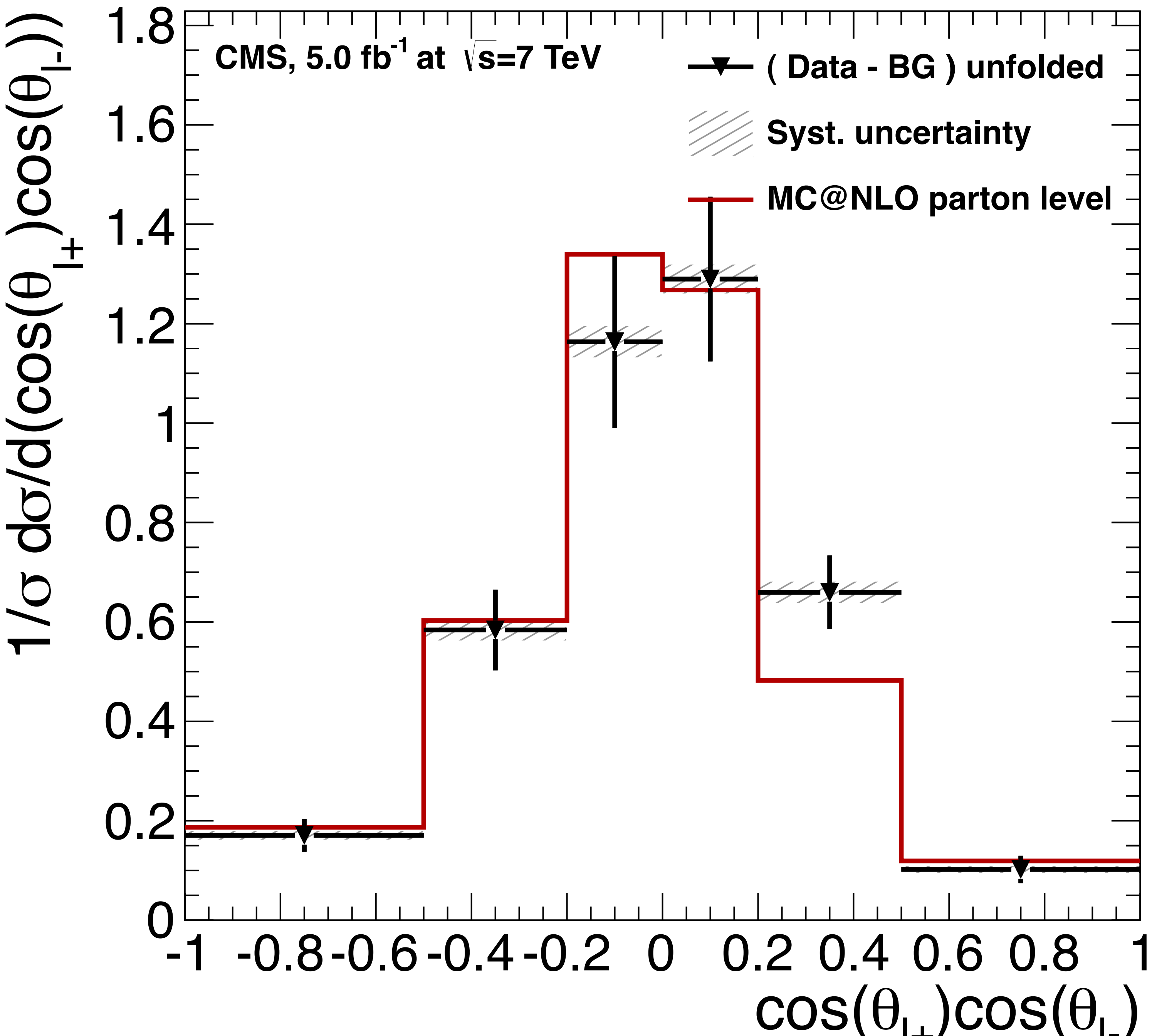
TUnfold, tau = 10



A_P=-0.018563I +/- 0.0149615

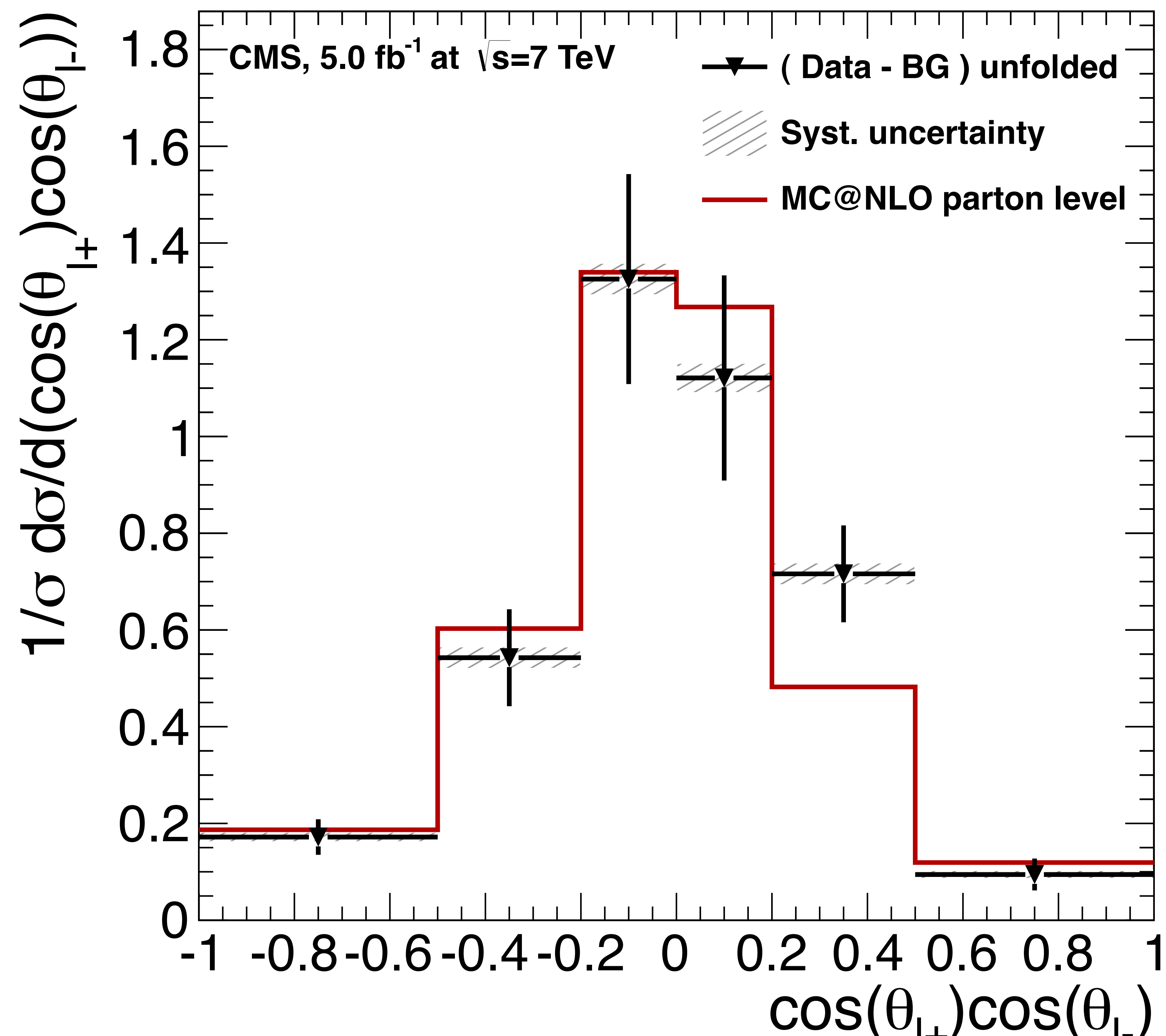
A_{c1c2} results

SVD, k=6



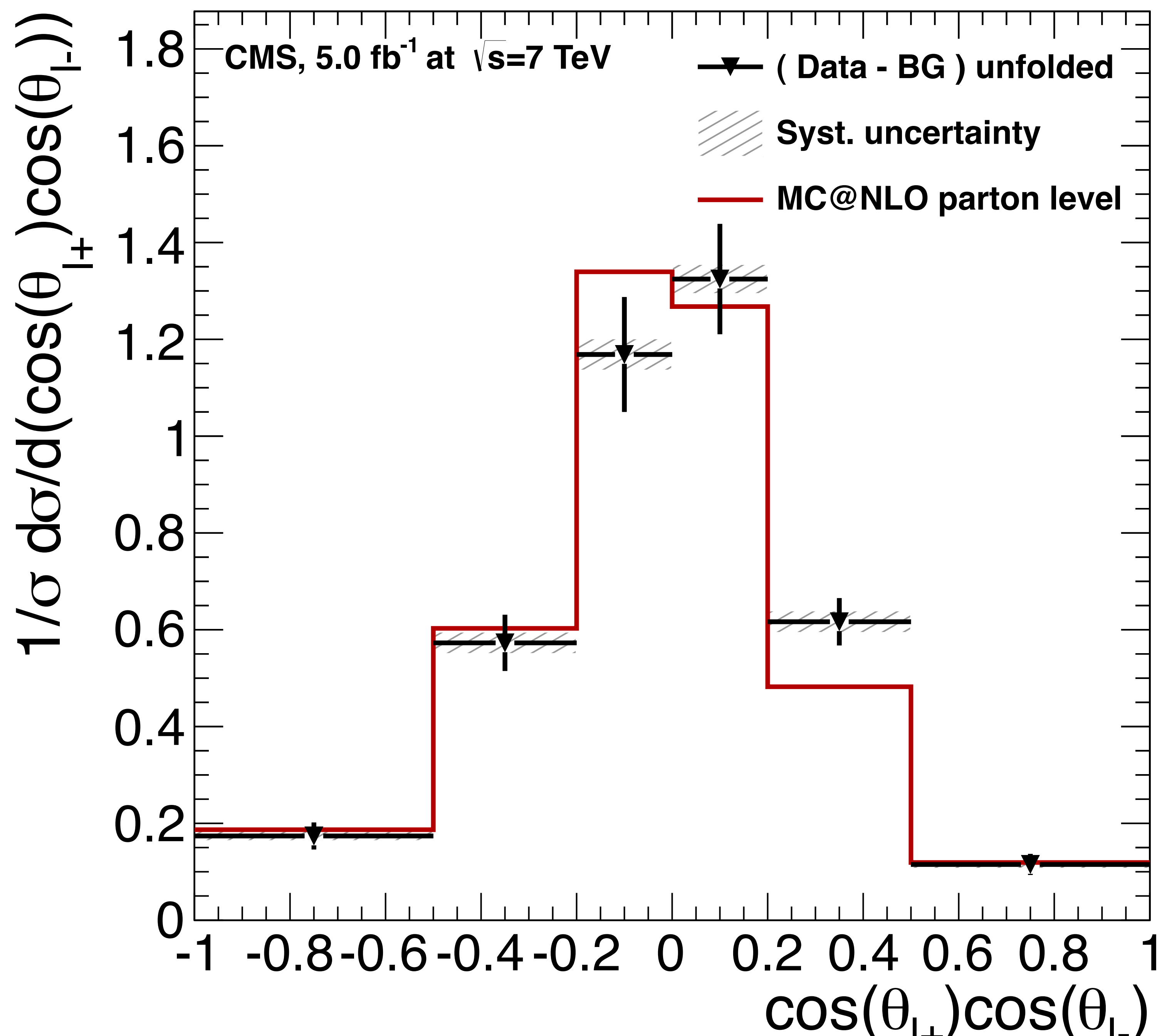
A_{c1c2}=0.0136347 +/- 0.0534609

TUnfold, tau = 0.0003



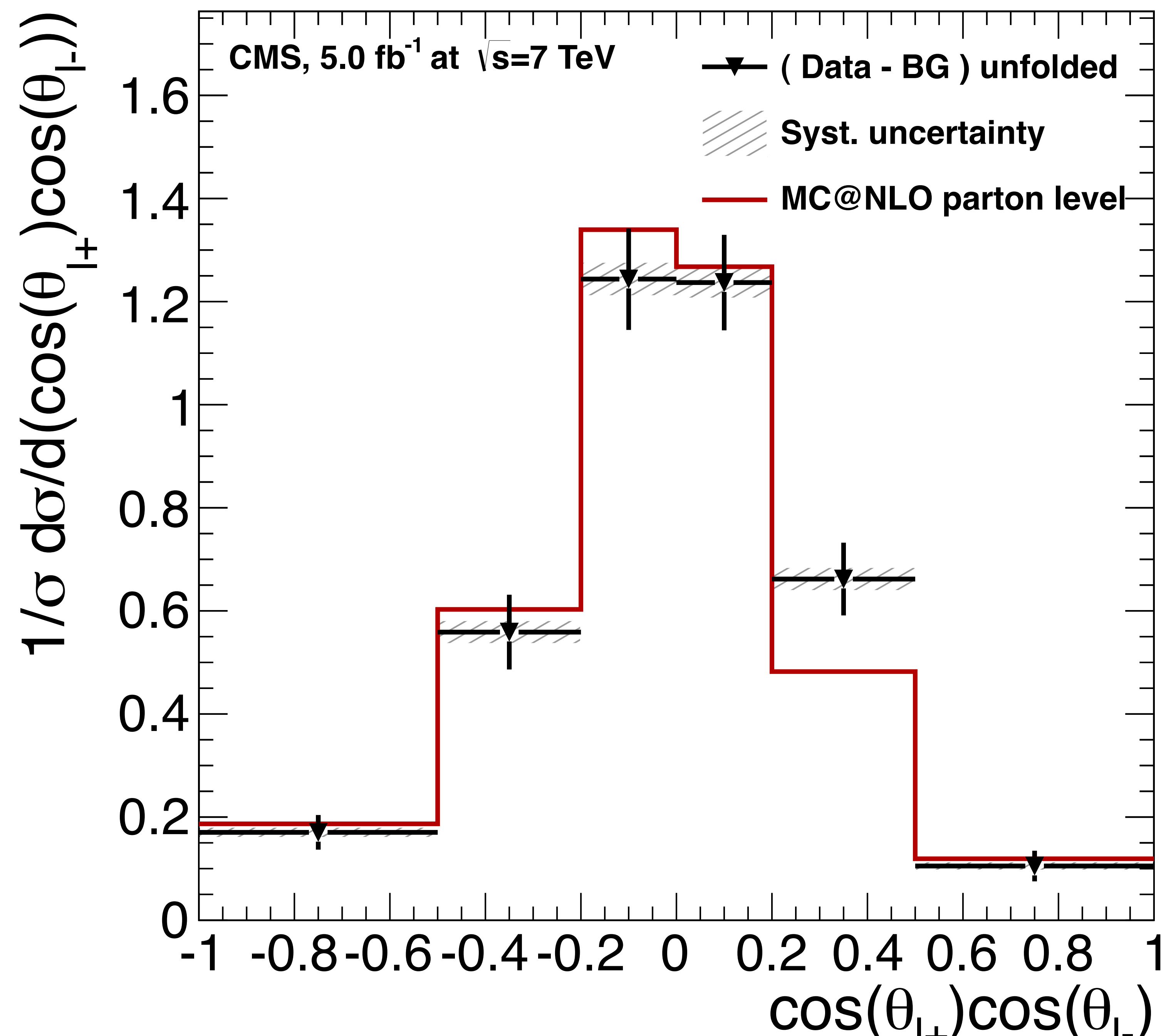
A_{c1c2}=-0.0276173 +/- 0.0620789

SVD, k=5



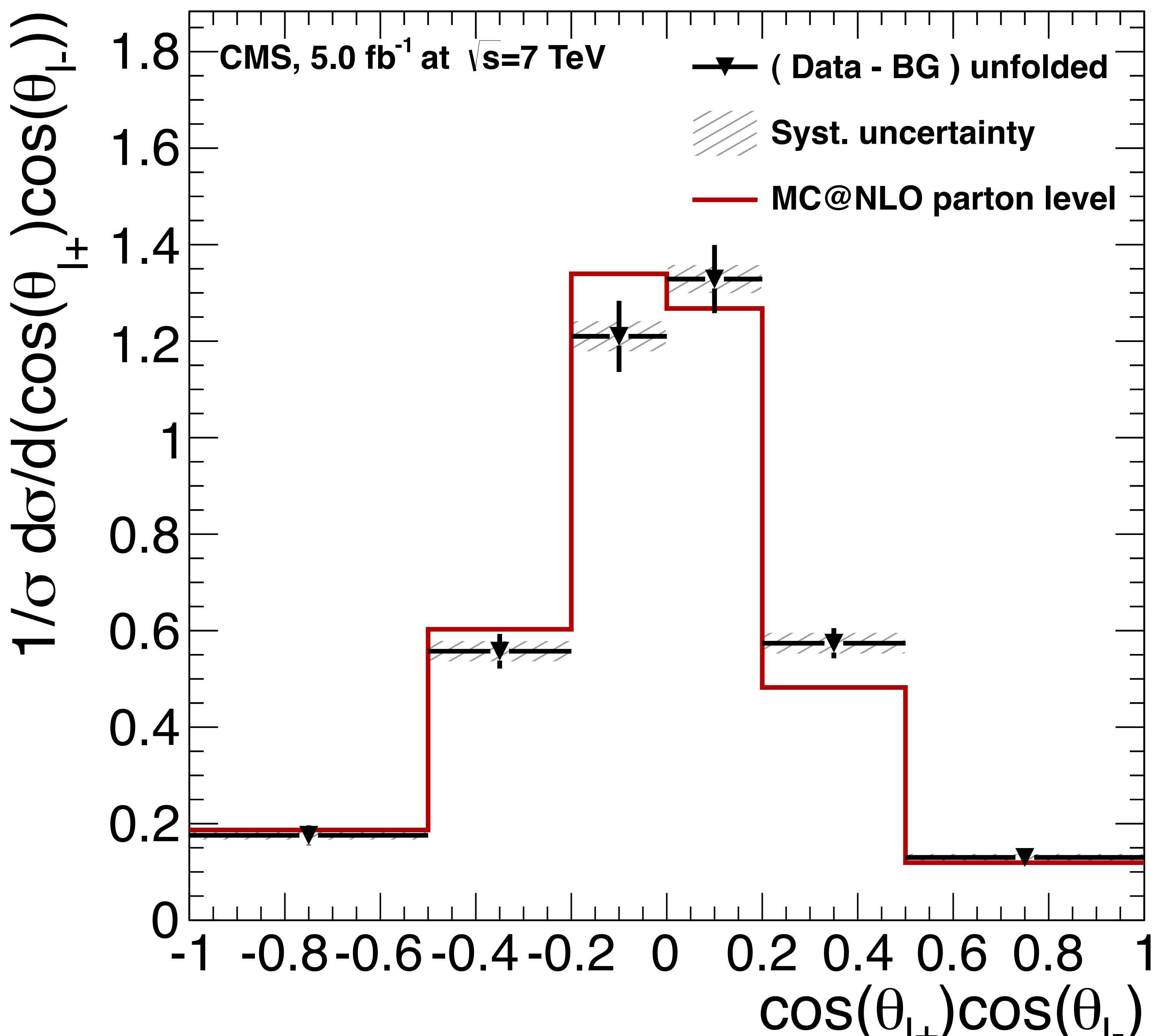
A_{c1c2}=0.0148291 +/- 0.0421044

TUnfold, tau = 0.001



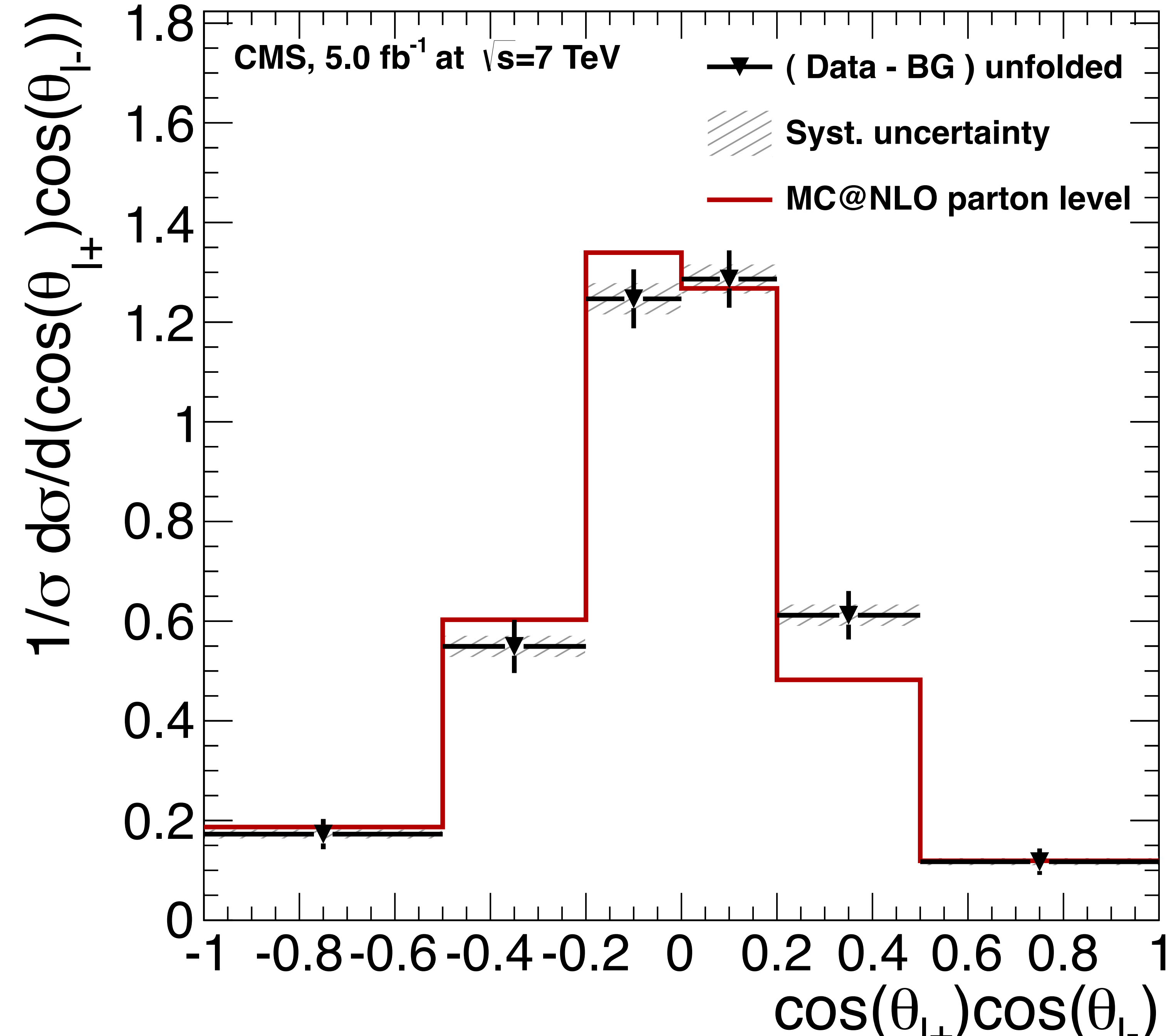
A_{c1c2}=-0.00310944 +/- 0.0355165

SVD, k=4



A_{c1c2}=0.00581684 +/- 0.0325673

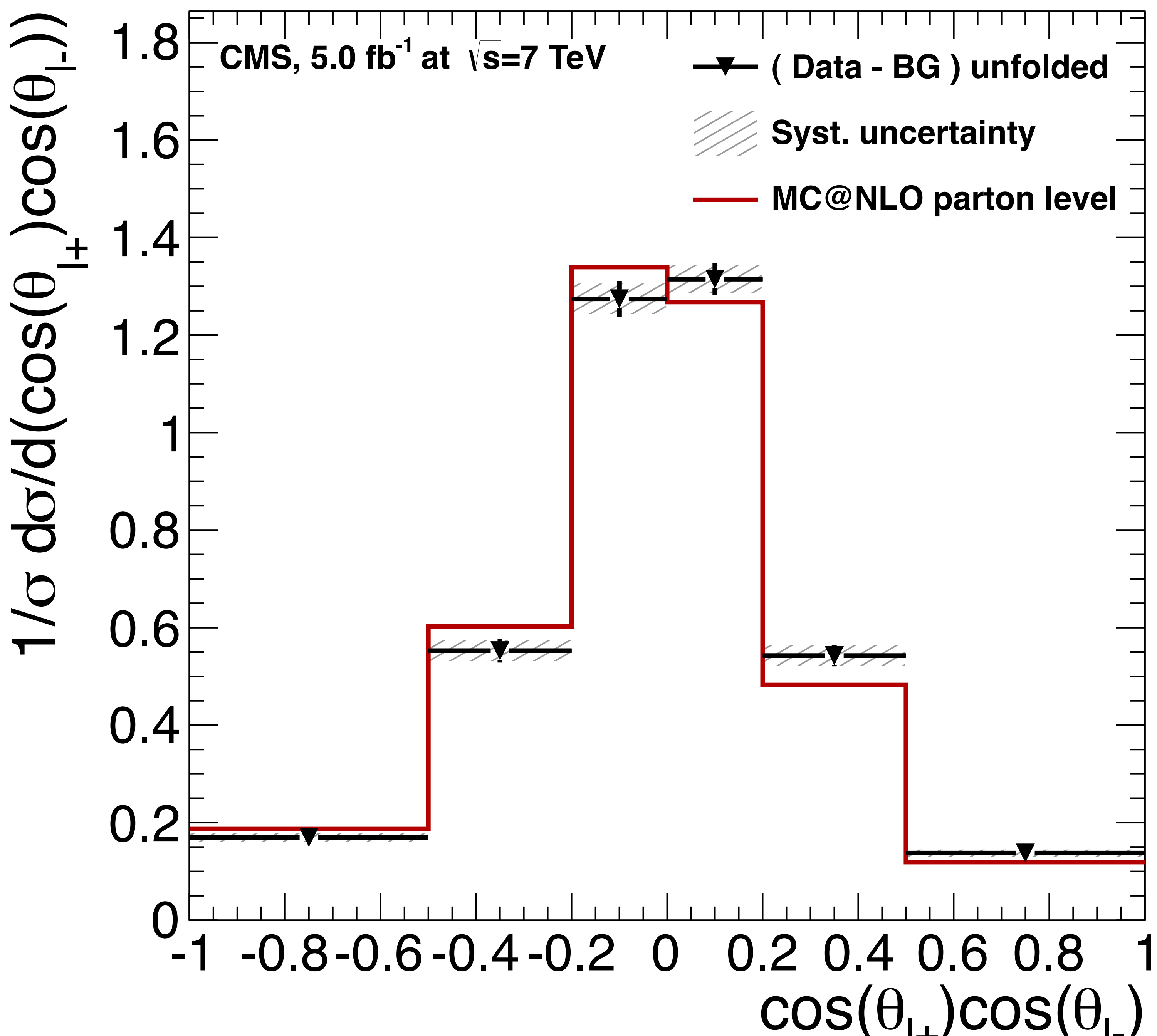
TUnfold, tau = 0.002



A_{c1c2}=-0.00094931 +/- 0.0291862

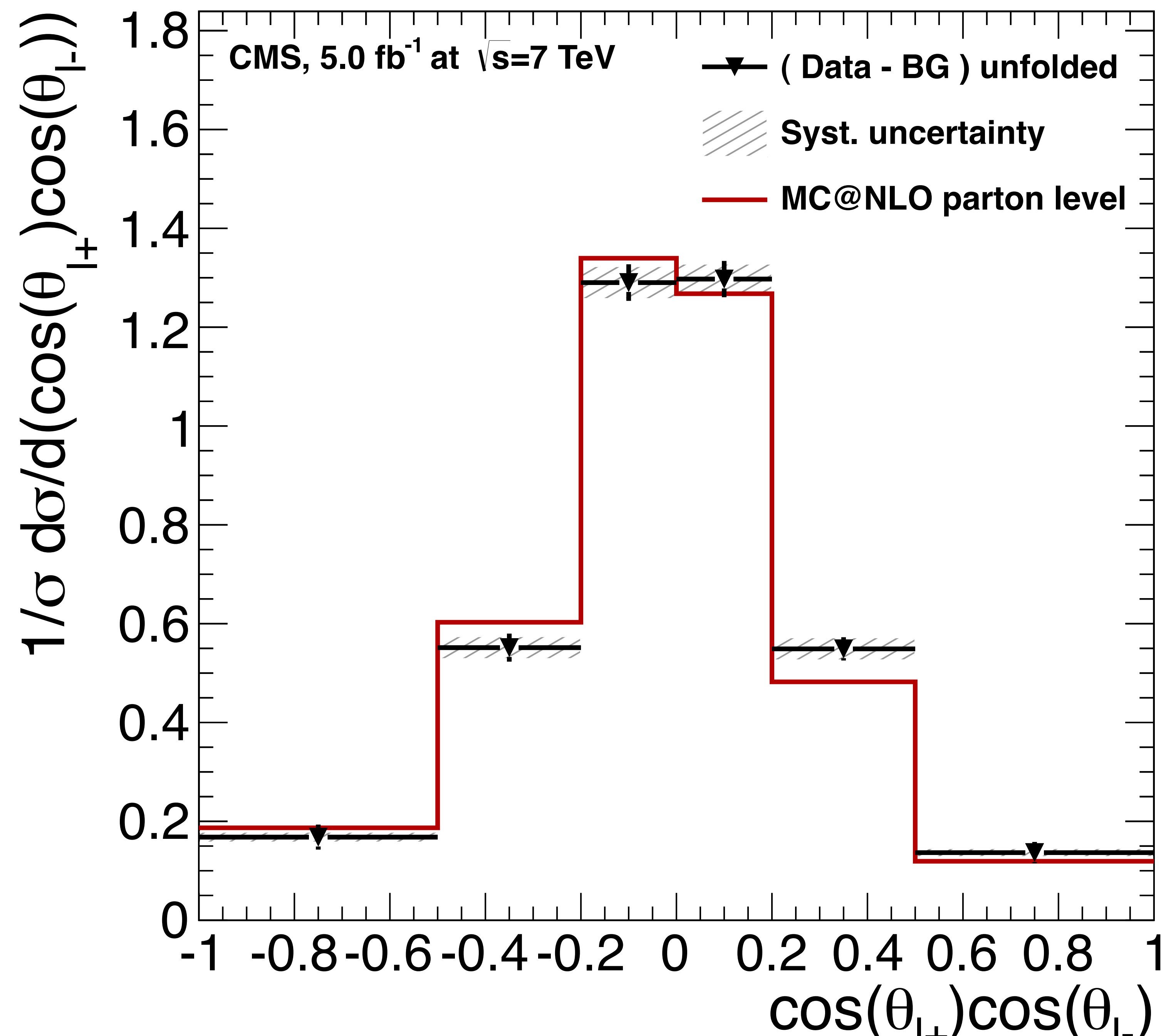
note, scanLcurve prefers tau ~ 0.005

SVD, k=3



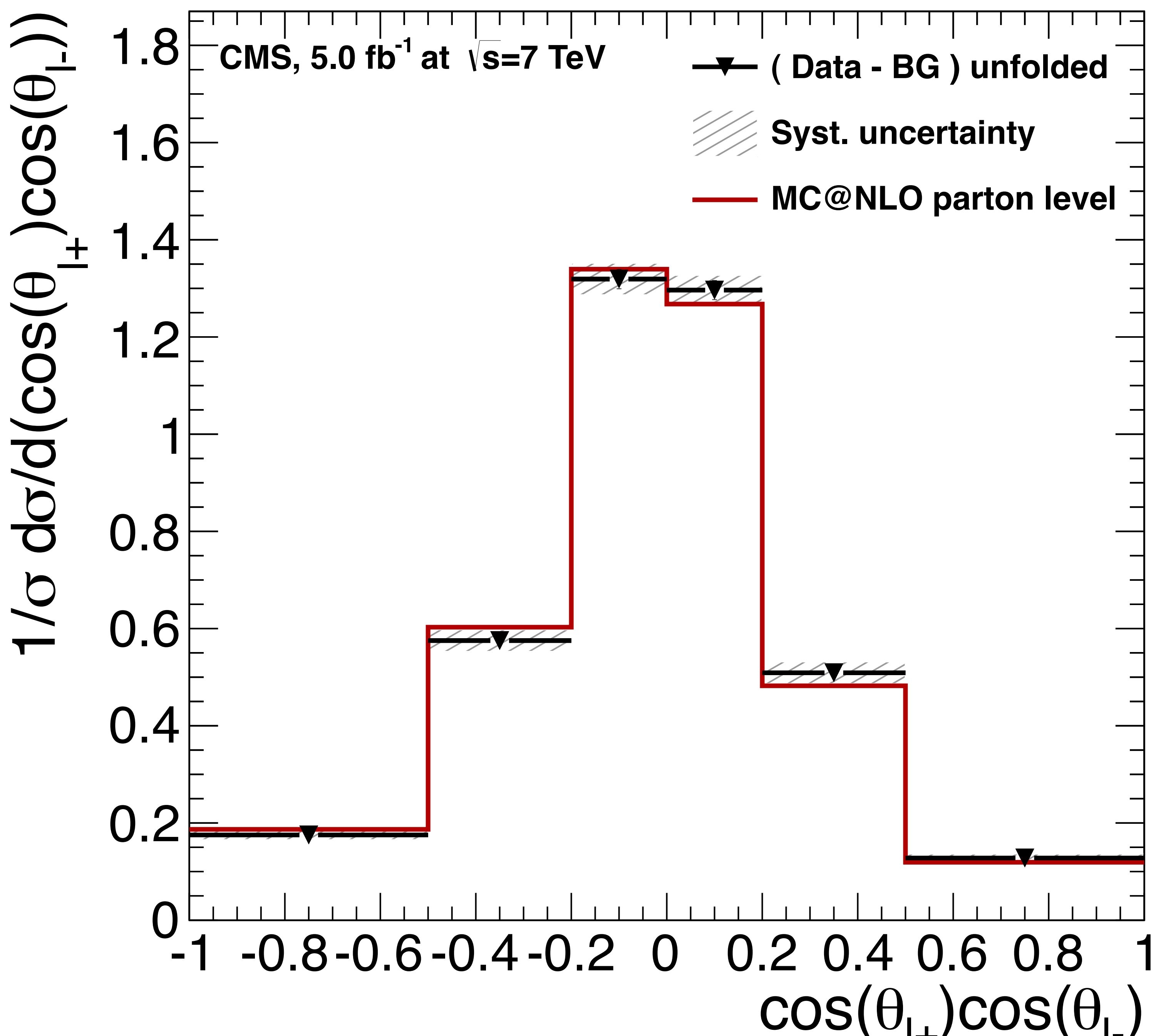
A_{c1c2}=-0.0109461 +/- 0.0265454

TUnfold, tau = 0.005



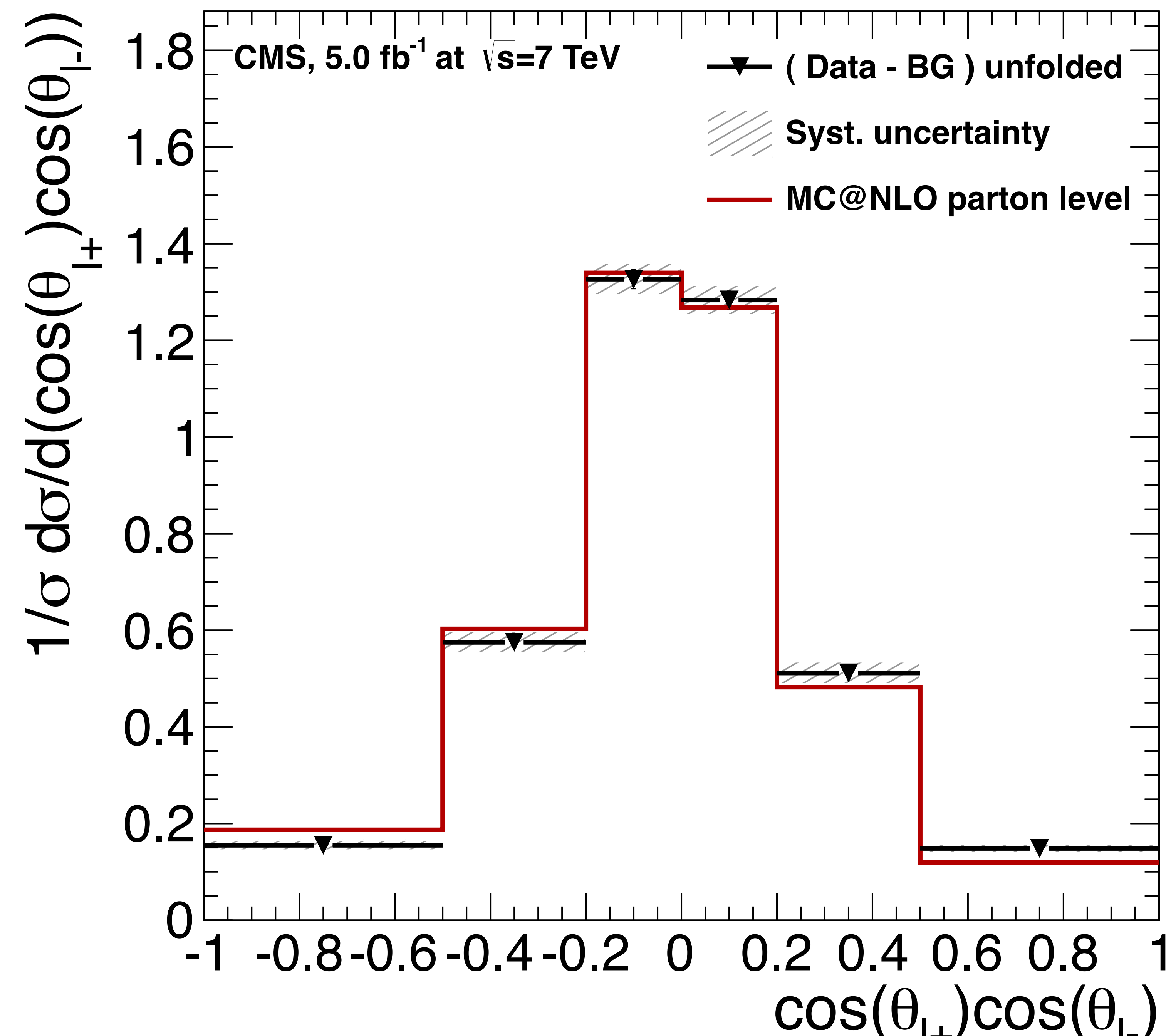
A_{c1c2}=-0.0150157 +/- 0.0232251

SVD, k=2



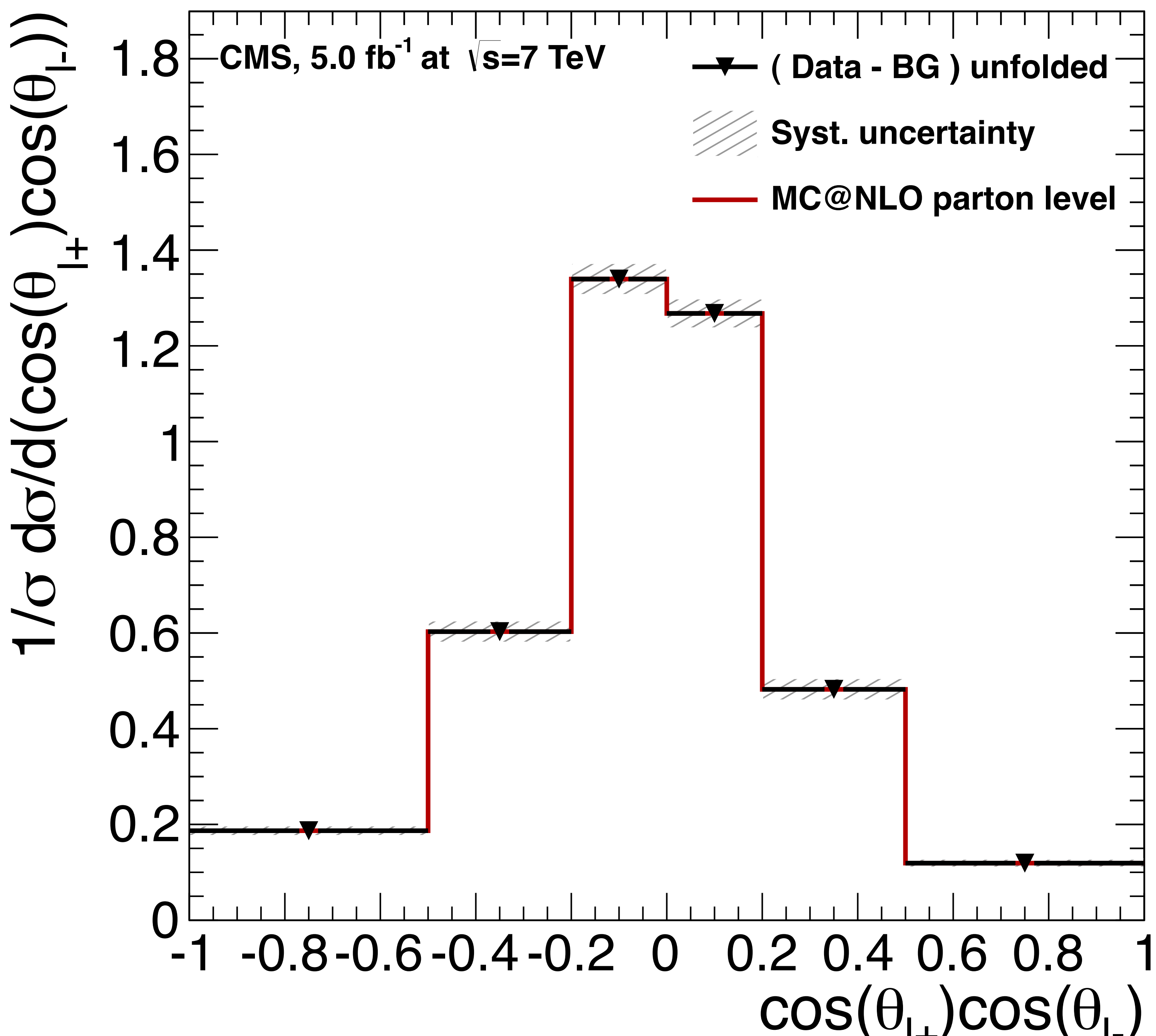
A_{c1c2}=-0.0481528 +/- 0.0136447

TUnfold, tau = 0.02



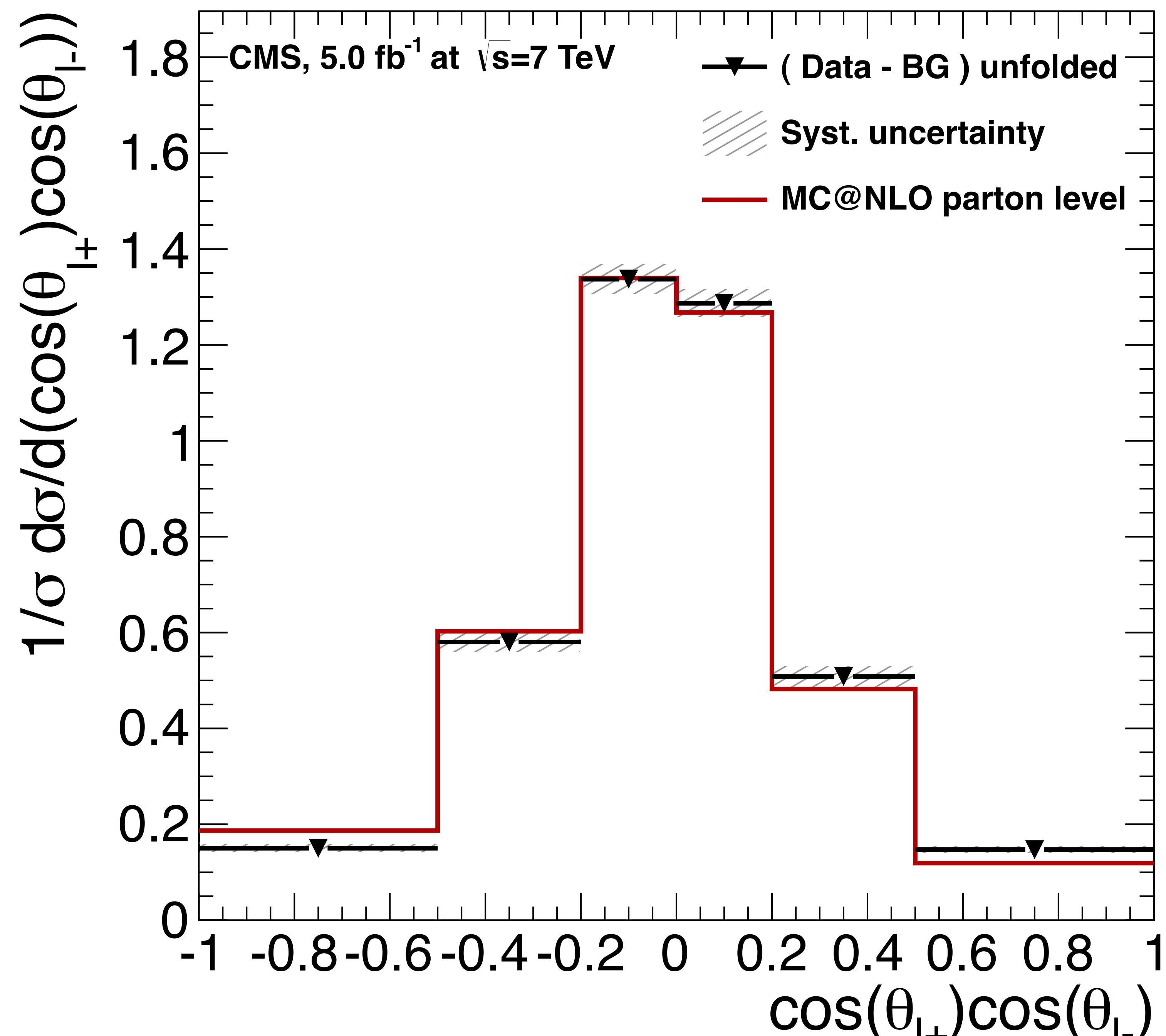
A_{c1c2}=-0.0309882 +/- 0.0204935

SVD, k=1



A_{c1c2}= -0.0844067

TUnfold, tau = 10



A_{c1c2}=-0.0332512 +/- 0.0204573

Summary

- TUnfold can reproduce SVD with $6 \leq k \leq 3$ very well
  - for the purely leptonic variables, even k=6 corresponds to relatively large tau, suggesting k=3 was over-regularising (we already knew this from our bin-by-bin linearity plots)
- The behaviour of the two methods is different in the limit of very strong regularisation (using SVD with k=1 the "unfolded" distribution exactly matches the MC truth, which doesn't happen in TUnfold with very large tau)
  - this doesn't affect us as we'd never use such strong regularisation
- On average, k=3 corresponds to tau~0.01, i.e. quite strong regularisation
  - this could explain why the TUnfold uncertainties from other groups are higher than our SVD uncertainties (they use less regularisation)
  - strong regularisation is not a problem as long as good linearity is maintained
    - we observed good linearity for the inclusive asymmetries, but not bin-bin
    - this effect is visible in the ~constant $A_{\Delta phi}$ but changing distribution shape as the regularisation is increased
- The useful range of tau (at least for 1D unfolding in 6 gen bins) is 0.0001 - 0.1, which matches what we saw with scanLcurve (decreasing or increasing tau beyond this range has little effect)

Conclusions

- RooUnfold SVD and TUnfold can produce very similar results, so no reason not to switch to TUnfold
- There were three reasons for the apparent discrepancy:
    - didn't apply strong enough regularisation in TUnfold (tau~0.01 is required, while tau~0.0001 is the default)
    - didn't set "bias scale" in TUnfold, so it defaulted to zero, meaning we had no bias distribution. Have to call DoUnfold(tau, hData_bkgSub, biasScale);
    - confusion due to bug that set the input data to reco-level MC for "pure TUnfold" but not "TUnfold via RooUnfold", which meant the two TUnfold implementations gave inconsistent results
- TUnfold in general has no problem with non-uniform bins, which were used throughout these slides (although it might adversely affect scanLcurve)
    - I found the results are also very similar when using the original 6 reco bins (advised against in the documentation, I think mostly because it breaks scanLcurve), instead of the 12 reco bins used in these slides (from splitting each of the 6 original bins in two)
- Since we'll be using TUnfold, we should continue working on how best to determine tau
    - tau~0.01 roughly matches the SVD results, but is probably over-regularising
    - we know the optimal range is ~ 0.001 - 0.01, based on these results and Dan's scanLcurve results, but can scanTau (TUnfoldDensity) be more precise than this?

Plan

- Implement TUnfoldDensity
  - find out which version of root is required and work with Terrence to get installed at UCSD
  - compare the results of scanTau and scanLcurve, and understand the two methods

Dan, with help from Jacob

- Run Linearity tests (inclusive and bin-by-bin) for TUnfold as implemented in these slides, using several values of tau (e.g. 0.002,0.005,0.01), and compare to SVD results in the AN

Jacob